

gene
CDS
14072. .14144,14960. .15376)
/gene="OSJNBa0014K08.2"
join(8072. .8086,9178. .9387,9474. .9671,12990. .13120,
14072. .14144,14960. .15376)
/gene="OSJNBa0014K08.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC05578.1"
/db_xref="GI:21902029"
/translation="MRSFLPVLPACGTGTPAPPPPTASLRRELALSPVLPARGC
VAGPNAATREAPAAVSSARVYVSPSLARRRTPPPPLFSPPVFSASP
FGLIPMRSLRLSGRRGSGDQAGALILSVADANYRGVLEQCEVEDQASFP
NPEVEVTEFFKDSGRISMNIEOALCMRFQKREHAROKLESYFOLCCTAAE
NGILLADDLAKPSRQYMEKSGATNRAIPDFRSAAASDAMRRARRSPCHANTR
NGRRGARDPEPSKQPIRVLGRLSDDELWVTSPPAAVFLRGETRAPAFALGTA
GRRGEDDSVLR
join(16191. .16218,17138. .17280)
/gene="OSJNBa0014K08.3"
join(16191. .16218,17138. .17280)
/gene="OSJNBa0014K08.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC05579.1"
/db_xref="GI:21902030"
/translation="MAITVEMNQPPSVGHCCIMRPSDYATCRVQACSCITAPGCT
GLREMAGLSSR"
join(19222. .19261,19694. .19848,19915. .20022,20628. .20832,
21203. .21300,21397. .21480,21572. .21649,21807. .21899,
21891. .22020,22277. .22421,22872. .22867,23073. .23252,
23347. .23453,23552. .23653,23956. .24011,24094. .24347,
24422. .24585,24909. .25092,25170. .25423)
/gene="OSJNBa0014K08.4"
join(19222. .19261,19694. .19848,19915. .20022,20628. .20832,
21203. .21300,21397. .21480,21572. .21649,21807. .21899,
21891. .22020,22277. .22421,22872. .22867,23073. .23252,
23347. .23453,23552. .23653,23956. .24011,24094. .24347,
24422. .24585,24909. .25092,25170. .25423)
/gene="OSJNBa0014K08.4"
/note="contains ESTs AU065754(s10005),AU175140(s10005)"
/codon_start=1
/product="putative kinase-like protein"
/protein_id="BAC05580.1"
/db_xref="GI:21902031"
/translation="MGTVNGEYEDPDANRAVITMLGGLPEFDLPDSDDELIRD
YLINGELCYADKLMGVLEGWGGVASDQSRNVKFLSVAAEMGLPGVGVLDLEG
SMSISIELCALKDVAATQLGHIINSTAKTPIRRKLELETDGPVLSVATPKRYPK
SOORSPILSGOKINEVVOFKGTITDLPARKISEMHSNLDNAPOTOSLIRVNGILD
ESTERKRGTEPHRVHLIRVIOETIERIGIOADHIRNOSIITKREDDYRSKIKALE
TIVNGTMEEMAINRELVYVKEKSIDEKRGCEDDMFLIREKNAIRITIASHOE
MQUANRHEQFRHDLKGLVDAENIRKVLNENKLFNEVQELKGNIRVORVPEPLRG
VLOGFLAIDYIGENGELTLANPSKQKEGRMKRFNKVFGTHSSQAEVSDIOLPLRS
VLGDFNVICFAYGOTGSGKTYTSGPOTSEDEGVNVRALNDLFISLKNAPFYES
NRAVGTALNERSRSRSHLIVHVRGLDYNVNGSTSGCLILIDLAGSERERSEADAG
RLEKAOHINKSLALGDVITSLAOKNAHVYRNSKLTLOVYOSGLGQAKIMVQNP
DISEYFETITLFAERVSGVELGAARNSGEGDITLFOVANSIDITARKMETIO
LOILKXSPMSKMDRNGSNLROSTISGTGSLPVAOQNOQISVTTITPLATQSLAG
SVEAEADNADSDGCSVGETEYSPAGASSETSAEAAVYLLFSTISCAKRSACDITGE
LMTVPRLRGT"
join(27978. .28537,28795. .29152,29378. .29650)
/gene="OSJNBa0014K08.5"
join(27978. .28537,28795. .29152,29378. .29650)
/gene="OSJNBa0014K08.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC05581.1"
/db_xref="GI:21902032"
/translation="MATASVYTLQPRELVPVEFRAPPPSPVTVGLTSSSGSLAG
ECPDNDDEIGRLRRSARVPLRLPERAVPRKRKAAMPVIVDVLALLPAGGPV
AEALRSAAVAFGCGVYGHGVDSIVSAALRHAAATARAATPEPEVEVNGDDENE

gene
CDS
35132. .35437)
/gene="OSJNBa0014K08.6"
complement(join(31081. .31442,31391. .31579,31787. .31815,
31903. .31969,32132. .32306,32410. .32549,32638. .32770,
33508. .33574,34086. .34243,34332. .34405,34496. .34609,
35132. .35437))
/gene="OSJNBa0014K08.6"
complement(join(31081. .31442,31391. .31579,31787. .31815,
31903. .31969,32132. .32306,32410. .32549,32638. .32770,
33508. .33574,34086. .34243,34332. .34405,34496. .34609,
35132. .35437))
/note="contains ESTs C72949(E2532),C72949(E2532)"
/codon_start=1
/product="putative casein kinase-like protein"
/protein_id="BAC05582.1"
/db_xref="GI:21902033"
/translation="MPVLSARARRAREQENPAAAAAAPPPARRRRARRRKEPEV
AVEAPEAEERKEEIEVADLREGGGKKEGDSGARSADKQAVDEGVTVPDT
VOVGSSPYRVADKILKGGFGQVYVRMLANGGASEVALKEFHTSGCNHGPPE
MOVYNAIGIGVPRVHKKRGQEVYVMTPHDYERNCFALSKILINLYERS
TKWRASGLHVVDQDPDPRFVEVEYVNLKEDPEYAKCISLEFTRVNPNDIP
INTDCAOKLILHVGKRGRLMEETDEQPKKIMGDPATQWISVYARPMKORHY
NVADSRLOHIDKNGEDLFTSCITSCSNMALLINDAGTGTSVYELSPHLEKWT
MDWERNYITILAGNAGSSLVMSKQTMTOOSKYVELDEFLPSGHIQRNDNG
RITAPAAWQDAAPVLSVRRPPIDEIETLRTSPAFSPCHVKSGHVEALSSRLLA
LVSQSAAYVYMLASECDLAVPYRAVAKIFENYRS"
join(39834. .39907,40580. .40722,41299. .41408)
/gene="OSJNBa0014K08.7"
join(39834. .39907,40580. .40722,41299. .41408)
/gene="OSJNBa0014K08.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAC05583.1"
/db_xref="GI:21902034"
/translation="MPYVGTGSCSAEYIRSLISROKEEYKRGKSELNPM
AVGSPIDAKSCFKHICDYSERILCVSIENRFARVSLDVQNRKRMVYVLCLODI
SLFFPC"
complement(join(47169. .47198,47300. .47377,47591. .47689,
47772. .47834,48987. .49058,49345. .49219,49310. .49366,
50328. .50531,50662. .50739,50830. .50892,50991. .51629,
51712. .51761,51841. .51921,52241. .52268))
/gene="OSJNBa0014K08.8"
complement(join(47169. .47198,47300. .47377,47591. .47689,
47772. .47834,48987. .49058,49345. .49219,49310. .49366,
50328. .50531,50662. .50739,50830. .50892,50991. .51629,
51712. .51761,51841. .51921,52241. .52268))
/gene="OSJNBa0014K08.8"
/note="hypothetical protein
similar to Arabidopsis thaliana chromosome 2, At2g19950"
/codon_start=1
/protein_id="BAC05584.1"
/db_xref="GI:21902035"
Query Match 81.6% Score 20.4; DB 8; Length 177263;
Best Local Similarity 95.5% Pred. No. 2,1e+02;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 ATATATCTACATCAACA 24
|||||
Db 92833 ATATATCTACATCAACA 92812
|||||
RESULT 10
AC115717 226726 bp DNA linear HTG 26-JUN-2002
LOCUS
DEFINITION Mus musculus clone RP23-28P10, WORKING DRAFT SEQUENCE, 9 ordered
pieces.
AC115717
AC115717.3 GI:21592233
AC115717.3

Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCACAAA 22
 |||
 Db 38714 CCTATATATCTACATTCACAAA 38735

RESULT 8
 AL137861/c 127682 bp DNA linear PRI 01-NOV-2000
 LOCUS Human DNA sequence from clone RP4-813B7 on chromosome 1 contains a
 DEFINITION pseudogene similar to zinc finger protein ZNF186, STS and GSSs,
 complete sequence.

ACCESSION AL137861
 VERSION AL137861.5 GI:9187172
 KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 127682)
 AUTHORS Cobley V.
 TITLE Direct Submission
 JOURNAL Submitted (18-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk

COMMENT

On Jul 14, 2000 this sequence version replaced gi:8694204.
 During difference assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.

The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1

RP4-813B7 is from the library RPCI-4 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-813B7.

FEATURES
 source location/Qualifiers

1..127682
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/chromosome="1"
 /clone="RP4-813B7"
 /clone_1lb="RPCI-4"

1..235
 /note="Alu repeat: matches 75..310 of consensus
 Alu repeat: matches 75..310 of consensus"

repeat_region 239..538
 /note="LIM4 repeat: matches 2107..2438 of consensus"

repeat_region 336..538
 /note="LIM4 repeat: matches 2227..2438 of consensus"

repeat_region 550..1553
 /note="LIM4 repeat: matches 2582..3669 of consensus"

misc_feature complement(565..1283)
 /note="match: GSS: Em:AQ383567"

repeat_region 675..1153
 /note="LIM4 repeat: matches 2732..3226 of consensus"

misc_feature

complement(698..1442)
 /note="match: GSS: Em:AQ896860"

misc_feature complement(906..1442)
 /note="match: GSS: Em:AQ761517"

misc_feature 1288..1539
 /note="match: GSS: Em:AQ879547"

misc_feature 1290..1693
 /note="match: GSS: Em:AQ826107"

misc_feature 1294..2055
 /note="match: GSS: Em:AQ901298"

misc_feature 1294..1860
 /note="match: GSS: Em:AQ47100"

misc_feature 1362..2055
 /note="match: GSS: Em:AQ490007"

misc_feature 1362..1961
 /note="match: GSS: Em:AQ585548"

misc_feature 1362..1761
 /note="match: GSS: Em:AQ275646"

misc_feature 1447..1837
 /note="match: GSS: Em:AQ355624"

misc_feature 1448..1964
 /note="match: GSS: Em:AQ474420"

misc_feature 1449..2054
 /note="match: GSS: Em:B76221"

misc_feature 1450..2017
 /note="match: GSS: Em:AQ537460"

misc_feature 1451..2223
 /note="match: GSS: Em:AQ38677"

misc_feature 1451..2150
 /note="match: GSS: Em:AQ38952"

misc_feature 1451..1896
 /note="match: GSS: Em:AQ669827"

misc_feature 1451..1877
 /note="match: GSS: Em:AQ810603"

misc_feature 1451..1827
 /note="match: GSS: Em:AQ18640"

misc_feature 1466..1877
 /note="match: GSS: Em:AQ692785"

repeat_region 1592..2208
 /note="LIMB1 repeat: matches 5338..5979 of consensus
 LIMB1 repeat: matches 5338..5979 of consensus"

repeat_region 2199..2404
 /note="LIM4 repeat: matches 3935..4135 of consensus
 LIM4 repeat: matches 3935..4135 of consensus"

misc_feature complement(2626..3066)
 /note="match: GSS: Em:AQ245895"

misc_feature complement(2887..3061)
 /note="match: GSS: Em:B91023"

misc_feature 3067..3302
 /note="match: GSS: Em:AQ009565"

misc_feature 3074..3453
 /note="match: GSS: Em:AQ261089"

misc_feature 3074..3302
 /note="match: GSS: Em:AQ059372"

repeat_region 3407..3448
 /note="match: GSS: Em:AQ059372"

repeat_region 3407..3448
 /note="match: GSS: Em:AQ059372"

misc_feature complement(3620..4114)
 /note="match: GSS: Em:AQ117010"

misc_feature complement(3956..4111)
 /note="match: GSS: Em:AQ304953"

misc_feature 4136..4803
 /note="match: GSS: Em:AQ531595"

repeat_region 4636..4942
 /note="match: GSS: Em:AQ531595"

repeat_region 5143..5591
 /note="match: GSS: Em:AQ224449"

misc_feature 5147..5738
 /note="match: GSS: Em:B60185"

misc_feature 5147..5659
 /note="match: GSS: Em:AQ076436"

misc_feature 5154..5605
 /note="match: GSS: Em:AQ608022"

/db_xref="taxon:4896"
 /chromosome="11"
 /clone="p1_p887"
 complement(join(1572..1631,1701..1836,1885..1997))
 /gene="SPBP8B7.01c"
 complement(join(1572..1631,1701..1836,1885..1997))
 /gene="SPBP8B7.01c"
 /note="SPBP8B7.01c, len:102"
 /codon_start=1
 /label="SPBP8B7.01c"
 /product="hypothetical protein"
 /protein_id="CAA21786.1"
 /db_xref="GI:3810825"
 /db_xref="SPTRMBL:094251"
 /translation="MSVERIKRAPAFRRKGIIIRKTHLIAMIRMRKKLDIDGSILOY
 ATGMAIPALVELSDKRYKROGYNISMDIATVEFMDEIVDEVSIONKRLGNVLLIRVSR
 "
 complement(1632..1646)
 /gene="SPBP8B7.01c"
 /note="ctaacatagctgttag, splice branch and acceptor"
 complement(1695..1700)
 /gene="SPBP8B7.01c"
 /note="gtacgc, splice donor sequence"
 complement(1837..1852)
 /gene="SPBP8B7.01c"
 /note="ctaacacagagcttag, splice branch and acceptor"
 complement(1879..1884)
 /gene="SPBP8B7.01c"
 /note="gtacgt, splice donor sequence"
 2917..3702
 /gene="SPBP8B7.02"
 2917..3702
 /gene="SPBP8B7.02"
 /note="SPBP8B7.02, len:261, SIMILARITY:Oxytlicha fallax,
 O 06905, hypothetical 129.9 kd protein.", (108 aa), fasta
 sco res: opt: 119, E():3.4e-05, (21.3% identity in 254
 aa)"
 /codon_start=1
 /label="SPBP8B7.02"
 /product="hypothetical protein."
 /protein_id="CAA21787.1"
 /db_xref="GI:3810826"
 /db_xref="SPTRMBL:094252"
 /translation="MNDESHISVLPVANKKOTSLVQLNKEETENOLKELEKRSQSLH
 KEEOQLQIYVAINDLRSTKEIKENKEEDLPFNFSASGDIANKKIQREDSRK
 BEESLHYKRYKVEKMKORSASKELEKLSAQQLQTVNVLRSNNDEILODMSEI
 KDCELEIQLQLOSASLINDPQNDRIAKKHSIDPNSRAHRSITNATNNYHIEF
 SLITTRHDSDEVSTVEKRTSKTISELEKLPADICIEDAFD"
 complement(4027..5118)
 /gene="SPBP8B7.03c"
 complement(4027..5118)
 /gene="SPBP8B7.03c"
 /note="SPBP8B7.03c, len:363"
 /codon_start=1
 /label="SPBP8B7.03c"
 /product="60s ribosomal protein 12"
 /protein_id="CAA21788.1"
 /db_xref="GI:3810827"
 /db_xref="SPTRMBL:094253"
 /translation="MAAARPTYSIKDGSVSEETIALPEVFKAPIPEDIVRSYTATA
 ANKKQPAYVSKACHQTSAESMGTRALARIPIRGGGGCTHRSGOAGANNCRSGO
 AFTKTRKHVAVNONEKRTAIISSAVAGSVPISLLANGHRIIEPVLPIYDDDVOS
 AKFPIKEAVALLEIKAYRDVNVANSKRLRAKGLKLRNRHVRGEPVLVNEDEGIL
 YLPENIIPGEIVNRRLLQLAPGSHGAGREVIWTKSAFGLDVSFGSTTAAQILKKR
 KAYKANVKLNTGKTPKAGGEKFLIVHEN"
 complement(4330..5064)
 /gene="SPBP8B7.03c"
 /note="Pfam match to entry PF00573 like, ribosomal protein
 1 family like"
 join(5912..5983,6026..8413)
 /gene="SPBP8B7.04"

```

CDS
join(5912, 5983,6026, .8413)
/gene="SPBP8B7.04"
/note="SPBP8B7.04, len:819"
/codon_start=1
/label="SPBP8B7.04"
/product="hypothetical protein"
/protein_id="CAA21789.1"
/db_xref="GI:3810828"
/db_xref="SPIRMBL:094254"
/translacion="MKNLANKILYNENDELIADMLLILKSPSSSDCLIDRDKM
ELVCKENLIGNERNPLVPLELPVNSLELQNSVNDMKRHYVSOVELNHTQVDF
TEQLPKRSKLNKNGTILRETSRMPSESTSETSNALDDSCFNDMMIIESEFSP
IDRGVTLAMPSONDILFTSKOLNNGGNFPEKATPEFLENMGISQAAATSTTPVCA
ISSVGLAINSGVDSNASHSSSTSELPCTGNTSITQVYENMGNSIKASPEYIAD
QOSKVSQONIKKEKPELPIPEPSLDEGFLNKNPTEPSNLYKSLSEKRSKWFTPL
VLRKPLVASSEFEETELKNTFTPLPLEALINHDCCFPDMLPSIKESSVAVPIP
KIKRSKSMFQSLQFNHCIPRSLSMINKOSQNLITTEITIRITCAALQVLEP
NILVILNKRPSSGSQLILASPTCTENIYIENRHLRFSKQSPQEQVISEFIDQVCK
VTEAKTTELSVILSERVLTATKLEPIKVKQNDNRCFKOLEYDIYILKRRLL
WELNLNPVIOLEIDYCHFKELKVNRLQIKNTYQTEPHLRPIKILQTSKHEY
FLHSESSYIFDAREDDYGVKLLKKNMISAAVYDNEANPECKDGTPTSTROOI
LLEELGIKREKRYLKEVRYVLDQPLPDPDSPLTVEPCOIIYVNTSLLEDSGYYSO
NGEFDREPNQNTTFDEKSERIILQNSTSTIGPAKNIKILLELIEVLNVNSTTAM
CLTLILGIVLILQVLEFYTN"
5984, 5983
/gene="SPBP8B7.04"
/note="gtaatt, splice donor sequence"
6013, 6025
/gene="SPBP8B7.04"
/note="ctaacccgttcgac, splice branch and acceptor"
complement(8688, .9422)
/gene="SPBP8B7.05c"
complement(8688, .9422)
/gene="SPBP8B7.05c"
len:244, SIMILARITY:porphyrinidum
purpureum, Q43060, carbonic anhydrase, (571 aa), fasta
scores: opt: 560, E():2.2e-31, (41.6% identity in 221 aa)"
/codon_start=1
/label="SPBP8B7.05c"
/product="carbonic anhydrase"
/protein_id="CAA21790.1"
/db_xref="GI:3810829"
/db_xref="SPIRMBL:094255"
/translacion="WPRDLKRRKIEDIDHDEITIDREASTSPVSGACKIDONGEIKRL
LERNLTWSQTSRKYPSFETAKIDIQTVQVLMWICSSRSEPTIILNLGEVFEVHN
IANVPRSDINALAVMEYSVLKAEKLIYVGHGCGGVAALPNNLNLDHLRII
ROVIDNREELDAIEDPOLRKLKAEHWTTRAQATSVTVRGVREAMEKRCLOVHGWIY
DLSNOIKRLDITDAIKKAKGYTDS"
complement(8754, .8822)
/gene="SPBP8B7.05c"
/note="Pfam match to entry PF00484 Pro_Ca,
prokaryotic-type carbonic anhydrases"
complement(8937, .9272)
/gene="SPBP8B7.05c"
/note="Pfam match to entry PF00484 Pro_Ca,
prokaryotic-type carbonic anhydrases"
join(10661, .10969,11033, .11056)
/gene="rpa2"
/note="SPBP8B7.06; rpl4-2"
join(10661, .10969,11033, .11056)
/gene="rpl4-2"
/note="SPBP8B7.06, len:110"
/codon_start=1
/label="rpa2"
/product="60S acidic ribosomal protein L4.2/L4B"
/protein_id="CAA21791.1"
/db_xref="GI:3810830"
/translacion="MKYLAAYLITVGGKSDPSADIESVLTGIEAESERTITLIN
EINGKRIDLIAAGNEKLATVPTGGAASAPAAAGAAPAAEAKAEAEAESESD
DMGFGLEFD"

```


Query Match
Best Local Similarity 91.7%; Score 20.8; DB 2; Length 152557;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

*	78100	80648:	contig of 2549 bp in length
*	80649	80748:	gap of unknown length
*	80749	83618:	contig of 2870 bp in length
*	83619	83718:	gap of unknown length
*	83719	86486:	contig of 2768 bp in length
*	86487	86586:	gap of unknown length
*	86587	89209:	contig of 2623 bp in length
*	89210	89309:	gap of unknown length
*	89310	91972:	contig of 2663 bp in length
*	91973	92072:	gap of unknown length
*	92073	94748:	contig of 2676 bp in length
*	94749	94848:	gap of unknown length
*	94849	97725:	contig of 2878 bp in length

QY 2 CATATATACCTACATCAAAACAA 25
|||||

Db 87874 CATATATATACATCAAAACAA 87897

RESULT 6
AC092723 188458 bp DNA linear PRI 06-FEB-2002
LOCUS Homo sapiens chromosome 16 clone RP11-542M13, complete sequence.
DEFINITION AC092723 AC068615
ACCESSION AC092723.3 GI:18542980
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 188458)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 188458)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 188458)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 6, 2002 this sequence version replaced g1:15055296.
Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES

Source
Location/Qualifiers
1..188458
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="16"
/clone="RP11-542M13".
BASE COUNT 49597 a 48044 c 45725 g 45092 t
ORIGIN

Query Match
Best Local Similarity 91.7%; Score 20.8; DB 9; Length 188458;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATATATACCTACATCAAAACAA 25
|||||

Db 65620 CATATATACCTACATCAAAACAA 65643

RESULT 7
SPBP8B7 71000 bp DNA linear PLN 15-FEB-2000
LOCUS S. pombe chromosome II pl p8B7.
DEFINITION AL032684
ACCESSION AL032684.1 GI:3810824
VERSION
KEYWORDS 60S ribosomal protein L4.2/L4B; autophagy; carbonic anhydrase; cell division control; transcriptional activator; cyclin; dbp2; DNA polymerase epsilon subunit b; ER lumen protein retaining receptor; LTR; NOL1 family; NOP2 family; p68-like protein; para-aminobenzoate synthase; peptidyl-prolyl cis-trans isomerase; puc1; RNA-binding; rpa2; rpl4-2; sun family; thiamine biosynthesis; tRNA nuclear export; ubiquitin carboxyl-terminal hydrolase; ubiquitin-protein ligase; zinc-finger.
Schizosaccharomyces pombe.
Schizosaccharomyces pombe.
Schizosaccharomycetes
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces.

SOURCE
ORGANISM Schizosaccharomyces pombe.
REFERENCE 1 (bases 1 to 71000)
AUTHORS Beck, A., Reinhardt, R., Lyne, M., Rajandream, M. A. and Barrell, B. G.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) European Schizosaccharomyces genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and Max Planck Institut fuer Molekulare Genetik, Innestr. 73, D-14195 Berlin, Germany

COMMENT

Notes:
Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.pombe/>)
During 1995 to 1996 about 66% of S. pombe chromosome I was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneMarker program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe). B (chromosome 2). c25H2 (cosmid name). .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
Pl p8B7 is overlapped at the 3' end by cosmid c19F5. EMBL entry SPBC19F5, accession number AL022599.

FEATURES

Source
Location/Qualifiers
1..71000
/organism="Schizosaccharomyces pombe"
/strain="972h-"

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1360: contig of 1360 bp in length
* 1461: gap of unknown length
* 1461: contig of 1050 bp in length
* 2511: gap of unknown length
* 2611: contig of 1111 bp in length
* 3722: gap of unknown length
* 3822: contig of 1109 bp in length
* 4931: gap of unknown length
* 5031: contig of 1304 bp in length
* 6335: gap of unknown length
* 6435: contig of 1088 bp in length
* 7523: gap of unknown length
* 7623: contig of 1247 bp in length
* 8870: gap of unknown length
* 8970: contig of 1334 bp in length
* 10304: gap of unknown length
* 10404: contig of 1202 bp in length
* 11606: gap of unknown length
* 11706: contig of 1275 bp in length
* 12981: gap of unknown length
* 13081: contig of 1083 bp in length
* 14164: gap of unknown length
* 14264: contig of 1552 bp in length
* 15816: gap of unknown length
* 15916: contig of 1826 bp in length
* 17742: gap of unknown length
* 17842: contig of 1734 bp in length
* 19576: gap of unknown length
* 19676: contig of 1192 bp in length
* 20868: gap of unknown length
* 20968: contig of 1619 bp in length
* 22587: gap of unknown length
* 22687: contig of 1995 bp in length
* 24682: gap of unknown length
* 24782: contig of 1294 bp in length
* 26075: gap of unknown length
* 26175: contig of 2012 bp in length
* 28187: gap of unknown length
* 28287: contig of 1004 bp in length
* 29292: gap of unknown length
* 29392: contig of 1926 bp in length
* 31318: gap of unknown length
* 31417: contig of 1543 bp in length
* 32860: gap of unknown length
* 32961: contig of 1453 bp in length
* 33061: gap of unknown length
* 34513: contig of 1603 bp in length
* 34613: gap of unknown length
* 36216: gap of unknown length
* 36316: contig of 2154 bp in length
* 38470: gap of unknown length
* 38471: contig of 2348 bp in length
* 38571: gap of unknown length
* 40919: contig of 2384 bp in length
* 41019: gap of unknown length
* 43403: contig of 2800 bp in length
* 43503: gap of unknown length
* 46403: contig of 2209 bp in length
* 46503: gap of unknown length
* 48711: contig of 2760 bp in length
* 48812: gap of unknown length
* 51571: contig of 1910 bp in length
* 51672: gap of unknown length
* 53581: contig of 1921 bp in length
* 53682: gap of unknown length
* 55602: contig of 1236 bp in length
* 55702: gap of unknown length
* 56938: contig of 1336 bp in length
* 57038: gap of unknown length
* 57090: contig of 2052 bp in length
* 59190: gap of unknown length

```

```

* 59191 62229: contig of 3039 bp in length
* 62230 62238: gap of unknown length
* 62330 64369: contig of 2240 bp in length
* 64370 64659: gap of unknown length
* 64670 67986: contig of 3317 bp in length
* 67987 68086: gap of unknown length
* 68087 71769: contig of 3683 bp in length
* 71770 71869: gap of unknown length
* 71870 74687: contig of 2828 bp in length
* 74698 74797: gap of unknown length
* 74798 77951: contig of 3154 bp in length
* 77952 78051: gap of unknown length
* 78052 83651: contig of 5610 bp in length.
* location/Qualifiers
  1..83651
  /organism="Rattus norvegicus"
  /db_xref="taxon:10116"
  /clone="CH230-118H3"
BASE COUNT 20702 a 18794 c 18175 g 20853 t 5137 others
ORIGIN
Query Match 83.2% Score 20.8; DB 2; Length 83651;
Best Local Similarity 91.7%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CATATATATCTACATTCMAAACNA 25
Db 30157 CATATATATCTCATCCMAAACNA 30180
RESULT 5
AC105540 152557 bp DNA linear HTG 13-JUL-2002
LOCUS Rattus norvegicus clone CH230-232018, *** SEQUENCING IN PROGRESS
DEFINITION ***70 unordered pieces.
AC105540.2 GI:21736405
VERSION HTG; HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 152557)
Mazny/D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralungu,H.C., Aze,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denu,A.L., Ding,Y., Dihn,K.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotte,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvan,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
L.J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louleaged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Meshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Messey,E., Mewney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,N., Morris,S.,
Mosier,M., Neal,D., Newton,U., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwoum,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

```

----- Project Information -----
 Center project name: H_NH1286011

 Summary Statistics
 Sequencing vector: M13; 28%
 Chemistry: Dye-terminator; Big Dye; 100% of reads
 Assembly program: Phrap; version 0.90319
 Consensus quality: 206799 bases at least Q40
 Consensus quality: 207057 bases at least Q30
 Consensus quality: 207288 bases at least Q20
 Insert size: 217000; agarose-ef
 Insert size: 207414; sum-of-ctgigs
 Quality coverage: 10.57 in Q20 bases; agarose-ef
 Quality coverage: 11.06 in Q20 bases; sum-of-ctgigs

 NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1076: contig of 1076 bp in length
 * 1077 1176: gap of unknown length
 * 1177 2354: contig of 1178 bp in length
 * 2355 2454: gap of unknown length
 * 2455 207614: contig of 205160 bp in length.
 * Location/Qualifiers
 1. 207614
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="UNK"
 /clone="RP11-1296011"
 1. 1076
 /note="assembly_name:Contig21"
 1177. 2354
 /note="assembly_name:Contig23"
 2455. 207614
 /note="assembly_name:Contig27
 clone_end:SP6
 vector_side:right"
 BASE COUNT 62251 a 42536 c 43366 g 59199 t 202 others
 ORIGIN
 Query Match 85.6% Score 21.4; DB 2; Length 207614;
 Best Local Similarity 95.7%; Pred. No. 86;
 Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 3 ATATATCTCATCTCAAAACAA 25
 ||||||| |||||||
 Db 153000 ATATATCTCATCTCAAAACAA 153022
 RESULT 4
 AC123500 83661 bp DNA linear HTG 24-JUL-2002
 LOCUS AC123500
 DEFINITION Rattus norvegicus clone CH230-118H3, *** SEQUENCING IN PROGRESS
 *** 41 unordered pieces.
 AC123500
 VERSION AC123500.2 GI:21909244
 KEYWORDS HTG; HTGS-PHASE1.
 SOURCE Rattus norvegicus.
 ORGANISM Rattus norvegicus;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 83661)
 Muray,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Albrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,

BOUCK,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dihn,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorell,J., H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Joliver,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kursth,A., Landy,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lightarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
 Lozado,R.J., Lu,X., Luchter,A., Luchter,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mayhew,E., McLeod,M.P., Martindale,A., Meier,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,S., Ogun,M., Okunolu,G.,
 Nguyen,N., Nickerson,E., Nwokenko,S., Ogun,M., Okunolu,G.,
 Otaguene,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,
 Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Tellrod,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczkyk,R., Wooten,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,S. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 83661)
 Worley,K.C.
 Direct Submission
 Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 83661)
 Worley,K.C.
 Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 19, 2002 this sequence version replaced gi:21240448.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GYHN
 Center clone name: CH230-118H3
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator; Big Dye; 100% of reads
 Assembly program: Phrap; version 0.90329
 Consensus quality: 43471 bases at least Q40
 Consensus quality: 46338 bases at least Q30
 Consensus quality: 48795 bases at least Q20

 NOTE: Estimated insert size may differ from sequence length
 * (see <http://www.hgsc.bcm.tmc.edu/docs/genbank/draft.data.html>).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 41 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

```

repeat_region      1. .71
/rpt_family="ERV1"
repeat_region      77. .1621
/rpt_family="L1"
repeat_region      1628. .1825
/rpt_family="L1"
repeat_region      2549. .2861
/rpt_family="Alu"
repeat_region      2835. .2871
/rpt_family="A-rich"
repeat_region      3021. .3191
/rpt_family="MER1-type"
repeat_region      4341. .4712
/rpt_family="MALR"
repeat_region      5872. .5907
/rpt_family="AT-rich"
repeat_region      5893. .6190
/rpt_family="Alu"
repeat_region      6458. .6754
/rpt_family="Alu"
repeat_region      7331. .7597
/rpt_family="L1"
repeat_region      7575. .7623
/rpt_family="L1"
repeat_region      7709. .8020
/rpt_family="AT-rich"
repeat_region      8162. .8500
/rpt_family="L1"
repeat_region      8671. .8915
/rpt_family="L1"
repeat_region      8931. .10971
/rpt_family="L1"
repeat_region      10998. .11270
/rpt_family="Alu"
repeat_region      11244. .11331
/rpt_family="(GAA)n"
repeat_region      11322. .12313
/rpt_family="L1"
repeat_region      12304. .12443
/rpt_family="(TA)n"
repeat_region      12408. .12474
/rpt_family="ERV1"
repeat_region      12479. .12824
/rpt_family="L1"
repeat_region      12825. .13431
/rpt_family="L2"
repeat_region      14960. .15290
/rpt_family="MALR"
repeat_region      15853. .15918
/rpt_family="Alu"
repeat_region      15919. .17117
/rpt_family="ERV1"
repeat_region      17123. .17365
/rpt_family="Alu"
repeat_region      17580. .17584
/rpt_family="Alu"
misc_feature       a126f06.sl"
17963. .18256
/Note="match to EST AA719796 (NID:g2732895) z954c06.sl"
repeat_region      18416. .18500
/rpt_family="MALR"
repeat_region      18501. .18810
/rpt_family="Alu"
repeat_region      18785. .18830
/rpt_family="(TAA)n"
repeat_region      18811. .19118
/rpt_family="MALR"
repeat_region      19397. .19634
/rpt_family="AT-rich"
repeat_region      20939. .20967
/rpt_family="AT-rich"
repeat_region      22281. .22302
/rpt_family="AT-rich"

```

```

repeat_region      22588. .22885
/rpt_family="Alu"
repeat_region      23047. .23095
/rpt_family="AT-rich"
repeat_region      23425. .23530
/rpt_family="L2"
repeat_region      23572. .24318
/rpt_family="MER2-type"
repeat_region      25329. .26141
/rpt_family="L1"
repeat_region      26142. .26436
/rpt_family="Alu"
misc_feature       26531. .26387
/Note="similar to Homo sapiens EST R36044 (NID:g792945)
y969e07.r1"
misc_feature       26536. .26368
/Note="similar to Homo sapiens EST AW881127
(NID:g8043137)"
repeat_region      26437. .27763
/rpt_family="L1"
repeat_region      28812. .29038
/rpt_family="L1"
repeat_region      29038. .29062
/rpt_family="(TTA)n"
repeat_region      29082. .29313
/rpt_family="Alu"
misc_feature       29114. .29147
/Note="similar to Homo sapiens EST AW881127
(NID:g8043137)"
repeat_region      29324. .29801
/rpt_family="L1"
misc_feature       29890. .30234
/Note="similar to Homo sapiens EST AW881127
(NID:g8043137)"

Query Match      85.6%; Score 21.4; DB 9; Length 175895;
Best Local Similarity 95.7%; Pred. No. 89;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 ATATATCTGACATTCGAAACAA 25
Db      28650 ATATATCTGACATTCGAAACAA 28672

RESULT 3
AC099800      207614 bp      DNA      linear      HTG 03-JAN-2002
LOCUS      Homo sapiens chromosome UNK clone RPL-1296011, WORKING DRAFT
DEFINITION      AC099800
ACCESSION      AC099800
VERSION      AC099800.2 GI:1804257
KEYWORDS      HTG; HTGS-PHASE1; HTGS-DRAFT; HTGS-FULLTOP.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 207614)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 207614)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (21-NOV-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT      On Jan 3, 2002 this sequence version replaced gi:17027321.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web Site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu

```

PDPEAQCVITITGFGDVARMLRRNCNHTFVMAISKMKPLHLEQIAHVRNPIONV
EMSEVDAPEKSGIATYTLRTAREKNOHLEIAKAKNELIFESAPENNNLHFTL
STYYVDVDEKSSSALMPEYERETITIIISPCYNNRNRNGIYIKYVGNIDVTL
ARRYIMDLPIISMCFNKNDMAPEPSVRNRMHIDEGCIIITKMPSPYEPADLS
GEVPLNCASLRKSEFNKKTLYAYOKYLSKFTPIAPQPDIDNSIWHHSIPNFKN
FNMPCRELSDGSGRRHRSSSIASSSKSYSKSKOSESSGGGSRSTRVSE
NSSVPIIMQPTPHFAPMLTPHHMLKYVYLQOHOAOQFLKGAAGLHGHIMPEP
PIYDGSFGARYSNNIYFSYIVQNNFNNGNHL"

gene
16033. 20426
/gene="T07F8.2"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T07F8.2;class
=Sequence"
join(16033. 16081,16586. 16697,16746. 16879,16933. 17322,
17739. 17866,18090. 18163,18212. 18319,20099. 20139,
20265. 20325,20372. 20426)
/gene="T07F8.2"
/note="exons 2 and 4 similar to transmembrane and
cytoplasmic domains of somatostatin receptor type 1"
/codon_start=1
/product="Hypothetical protein T07F8.2"
/protein_id="AA38097.1"
/db_xref="GI:1708759"
/translation="MPEIRREIRPQISRYEEMASPYFAGFLIATVLTGPNINW
CSVARSRKPKSLMSRSPSDRLRAYISLAVIDITVALLVRLHFLPHMLDMS
CRAMFVLENSVKITSLTLCISLERITIRKPESEVRQFNAPIDGSIQGLVY
GAIIVQINSVYSSDGLNCVRSYKKAIPKASVLAVALVLTITLSIVQIVRY
RRKTRRARQANSRVRESIVNEPRYEMTAIVAGVHVLVAVNPNRLNLISAV
TIOSELGARIIRFENFRDYSTRRLFIATIMTSMNAGMIVYAVNPNRLNLISAV
AKTYFSSATGEGKISSRAGSNINPNLQRTITFCFTFNSQSTRKL"
21220. 23732
/gene="T07F8.1"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T07F8.1;class
=Sequence"
join(21220. 23253,23298. 23732)
/gene="T07F8.1"
/note="weak similarity in 5' end of exon 1 to C. elegans
myosin heavy chain B; 3' end of exon 1 similar to ZK686.4"
/codon_start=1
/product="Hypothetical protein T07F8.1"

gene
21220. 23732
/gene="T07F8.1"
/note="for a graphical representation of this gene see:
http://www.wormbase.org/db/seq/sequence?name=T07F8.1;class
=Sequence"
join(21220. 23253,23298. 23732)
/gene="T07F8.1"
/note="weak similarity in 5' end of exon 1 to C. elegans
myosin heavy chain B; 3' end of exon 1 similar to ZK686.4"
/codon_start=1
/product="Hypothetical protein T07F8.1"

Query Match 85.6%; Score 21.4; DB 3; Length 23862;
Best Local Similarity 95.7%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CATATATATCTACATTCATAAACA 24
|||||
DB 14631 CATAAATCTACATTCATAAACA 14653

RESULT 2
AC064862 175895 bp DNA linear PRI 09-JAN-2002
LOCUS
DEFINITION Homo sapiens BAC clone RP11-368M16 from 7, complete sequence.
ACCESSION AC064862
VERSION AC064862.6 GI:15145584
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 175895)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 175895)
AUTHORS Kruchowski, S., Kozlowski, A., Hawkins, M., Elliot, G. and Doeber, A.
TITLE The sequence of Homo sapiens BAC clone RP11-368M16
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 175895)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Waterston, R.H.
Direct Submission
Submitted (23-APR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 175895)
Waterston, R.H.
Direct Submission
Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 175895)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced gi:14589722.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.edu
----- Summary Statistics -----
Center project name: H_NH0368M16

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 17 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR17, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J., and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://dpcpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP11-368M16;
Actual end is at base position 175895 of RP11-368M16.

FEATURES
Source

RP11-368M16 contains single plasmid regions from 51174 to 51278 and
80620 to 80630.
Location/Qualifiers
1. 175895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="7"
/clone="RP11-368M16"
/clone_id="RPCT-11"

JOURNAL Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613
 PUBMED 9851916
 REFERENCE 2 (bases 1 to 23862)
 AUTHORS Pauley, A.
 TITLE The sequence of *C. elegans* cosmid T07F8
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 23862)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (04-DEC-1996)
 REFERENCE 4 (bases 1 to 23862)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUL-2001) Department of Genetics, Washington University, Genome Sequencing Center, 444 Forest Park Avenue, St. Louis, MO 63110, USA
 REFERENCE 5 (bases 1 to 23862)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 444 Forest Park Avenue, St. Louis, MO 63110, USA

COMMENT

Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 0RO, England
 email: r.w@hema.mede.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

For a graphical representation of this cosmid sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T07F8;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is F10E7, 1000 bp overlap; the 3' cosmid is C27H5, 500 bp overlap. Actual start of this cosmid is at base position 1 of T07F8; actual end is at 12559 of C27H5.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yujl Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFeome cloning project (<http://wormid.dfci.harvard.edu/>), similarity to other proteins from Blast analysis (<http://blast.wustl.edu/>), sequence conservation with *C. briggsae* using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual *C. elegans* Genbank submissions, and personal communications with *C. elegans* researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

Source

1. 23862
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"

gene
 /chromosome="II"
 /clone="T07F8"
 complement(2668..3987)
 /gene="T07F8.4"

CDS
 /note="For a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=T07F8.4;class=Sequence>"
 complement(join(2668..2845,2887..2948,3001..3353,3401..3596,3646..3755,3798..3854,3906..3987))
 /gene="T07F8.4"

gene
 /note="Contains similarity to Pfam domain: PF00249 (myb_DNA-binding), Score=11.3, E-value=0.026, N=1; coded for by the following *C. elegans* cDNAs: YK132d3.5, YK132d3.3, YK479h9.3, YK479h9.5"
 /codon_start=1
 /product="Hypothetical protein T07F8.4"

protein_id="AA838099.1"
 /db_xref="GI:1708761"

gene
 /translation="MDEGSRSSITDEEDVDNDNGEDYEATIEEELLDEGDYDE
 LKLEDEGMSIELOKXGVRPRAESSANNEDTEAEENVGSSAAEPHDSTIFD
 MGNMSGFDDEDDDYAPPPPMKRSITVDVLCQADVPLENEATVESAREDTIAT
 IDONTNPSDEVINDYVKDYGKRAHQDPGATGSRDEEDLALYSGNPTKAK
 ESPPPHINAPFRVTRSDALGDESEKAPESSELEIKGDFSLIRRLPYKRVGELI
 EYVQKLTGPGYRVMDARHPQAPVYQPHLSAHHQVQALDQNGCTGFVASESP
 STSEPESTLTN"
 4614..9003
 /gene="T07F8.3"

gene

note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=T07F8.3a;class=Sequence>"
 join(4614..4738,4787..4868,4910..5020,5066..5136,5179..5344,5389..5515,5560..5681,5728..5805,5850..5987,6035..6277,6326..6424,6473..6641,6784..6906,6951..7019,7355..7432,7477..7716,7764..8190,8310..8591,8820..9003)
 /gene="T07F8.3"

CDS

note="coded for by the following *C. elegans* cDNAs: YK609d3.3, cm13g9, YK29c11.5, YK163e6.5, YK35f11.5, YK609d3.3"
 /codon_start=1
 /product="Hypothetical protein T07F8.3a"

protein_id="AAK7629.1"
 /db_xref="GI:15011798"

translation="MGEOSHEKDHDEASHNPVRSAVEDADTRQLQMAENASAKL
 FVSALADIVNENFTHDQSAQVADENQNRQVSYNTRKRAHILQKAGLPLPS
 NIPMIEINPRTVLNMEFEQYSLMTSDGHEVNASIAETNLIQLPDRSGVGT
 PDPEAQVITIGYFGVDYRAIMLRNCHETVMAISKMKMPLHEQAVRQVPIQNV
 EMGFVDAPEKNGIIVTILITAREKQHELEAKRLELIEFESPAENNFILHTL
 STYVVDGLSSSTQALMPYIERETITIIISPCYNNRNRGRIEIKVGNINDVLK
 ARYIMDLPIISMCPNKNMDMAPEPSRPNHMTIDSGILTKTPSYVEPADLIS
 GEVPLNASTRSKEPNKRIYKQVLSKPEFIPAPDPDYNSTIMHSLPRNFKN
 FNPCKRELSDGNGRRHRSSIASRSKSHYSKAKQFSESGGFSRSHTRYSSE
 NSSTVPTMCPPTPHFAPEMLTPHHMLKVVYLOHQAOQFLKGAAGLHGHIMPP
 PLIVDSFVAFALPPADPVYDFGPFYVHGGLFVVEADQHRNRRSSPLSGEIRKPS
 RMNGNRSSSTGYSYPTPRQRYVEQVDRDLRSHIGSRRTSVNCDQNVSMHOG
 YEROYPRQHLQKDDQKMTGSGDINSSRTINVHRYVNRNEDYDFVNGSPARK
 SPLEDOYLOLQMTHLKLSMDNDLCHETKIHMSPPNCDSTYDASGENDMDGQF
 QRLSNANINESSRRPRYTSCTFEKQGSARITDSGASVYDHASTHOSRSTDRK
 VGNNGVTKITLLEPRAREKDYKISLEHTKYSNEGEDSKSENSTLDSGRQYID
 PMKRNITILNRRNSFFYLHQFLSLYNNPDLAYRIRLHQFLSYSPSICLSRISLYP
 RLYDY"

CDS

join(4614..4738,4787..4868,4910..5020,5066..5136,5179..5344,5389..5515,5560..5681,5728..5805,5850..5987,6035..6277,6326..6424,6473..6641,6784..6906,6951..7108)
 /gene="T07F8.3"
 /note="coded for by the following *C. elegans* cDNAs: YK345b5.3, YK719f7.3, YK345b5.5, YK19f7.5"

codon_start=1
 /product="Hypothetical protein T07F8.3b"
 /protein_id="AAK7630.1"
 /db_xref="GI:15011799"

translation="MGEOSHEKDHDEASHNPVRSAVEDADTRQLQMAENASAKL
 FVSALADIVNENFTHDQSAQVADENQNRQVSYNTRKRAHILQKAGLPLPS
 NIPMIEINPRTVLNMEFEQYSLMTSDGHEVNASIAETNLIQLPDRSGVGT

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 88.7891 Seconds

(without alignments)
8194.363 Million cell updates/sec

Title: US-09-477-082-32

Perfect score: 25
Sequence: 1 ccatatcatctacattcaaaacaa 25

Scoring table: IDENTITY-NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_on:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21.4	85.6	23862	3	U80954	U80954 Caenorhabditis
2	21.4	85.6	175895	2	AC064862	AC064862 Homo sapi
3	21.4	85.6	207614	2	AC099800	AC099800 Homo sapi
4	20.8	83.2	83661	2	AC123500	AC123500 Rattus no
5	20.8	83.2	152557	2	AC105540	AC105540 Rattus no
6	20.8	83.2	188458	2	AC092273	AC092273 Homo sapi
7	20.4	81.6	71000	8	SPB887	AL032654 S. pombe c
8	20.4	81.6	127682	9	AL137861	AL137861 Human DNA
9	20.4	81.6	177283	8	AP003376	AP003376 Oryza sat
10	20.4	81.6	226726	2	AC115717	AC115717 Mus muscu
11	20.2	80.8	270987	2	AC122828	AC122828 Mus muscu
12	20.2	80.8	17280	6	AX251530	AX251530 Sequence
13	20.2	80.8	25539	9	AL161775	AL161775 Human DNA
14	20.2	80.8	114016	9	AL365396	AL365396 Human DNA
15	20.2	80.8	140137	2	AC027611	AC027611 Homo sapi
16	20.2	80.8	183786	2	AL807741	AL807741 Mus muscu
17	20.2	80.8	185237	2	AC111638	AC111638 Rattus no
18	20.2	80.8	252079	2	AC115349	AC115349 Rattus no
19	20.2	80.0	68726	8	AB017063	AB017063 Arabidops
20	20.0	80.0	222323	2	AC091254	AC091254 Mus muscu
21	19.8	79.2	5474	6	AX458550	AX458550 Sequence
22	19.8	79.2	46517	5	AL672060	AL672060 Zebrafish
23	19.8	79.2	103194	8	AC006223	AC006223 Arabidops
24	19.8	79.2	119958	2	AC104706	AC104706 Oryza sat
25	19.8	79.2	122680	2	AC130231	AC130231 Rattus no
26	19.8	79.2	123182	9	AC099339	AC099339 Homo sapi
27	19.8	79.2	129420	8	AC087547	AC087547 Oryza sat
28	19.8	79.2	141041	8	AC087550	AC087550 Oryza sat
29	19.8	79.2	156533	2	AC117070	AC117070 Dictyoste
30	19.8	79.2	156721	9	AL139389	AL139389 Human DNA
31	19.8	79.2	162410	2	AL133265	AL133265 Homo sapi
32	19.8	79.2	166978	10	AC127431	AC127431 Mus muscu
33	19.8	79.2	168182	2	AC023799	AC023799 Homo sapi
34	19.8	79.2	178815	2	AC100883	AC100883 Mus muscu
35	19.8	79.2	186044	9	AC009803	AC009803 Homo sapi
36	19.8	79.2	194005	2	AC130723	AC130723 Mus muscu
37	19.8	79.2	201500	2	AC103946	AC103946 Mus muscu
38	19.8	79.2	203335	2	AC122056	AC122056 Mus muscu
39	19.8	79.2	203713	2	AC084725	AC084725 Mus muscu
40	19.8	79.2	209320	2	AC016884	AC016884 Homo sapi
41	19.8	79.2	223940	2	AC087567	AC087567 Mus muscu
42	19.4	77.6	65869	2	AC113552	AC113552 Homo sapi
43	19.4	77.6	106866	2	AC126007	AC126007 Medicago
44	19.4	77.6	133181	9	AC100854	AC100854 Homo sapi
45	19.4	77.6	166114	9	AC083837	AC083837 Homo sapi

ALIGNMENTS

RESULT 1
U80954
LOCUS U80954
DEFINITION Caenorhabditis elegans cosmid 107f8, complete sequence.
ACCESSION U80954
VERSION U80954.1 GI:1708757
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Felodermidae; Caenorhabditis.
REFERENCE 1 (bases 1 to 23862)
AUTHORS Waterston, R.
TITLE Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium

Pred. No. is the number of results predicted by chance to have a

Best Local Similarity 89.5%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGTG 20
|||
Db 715 AGGGGATTGTAGATTGTG 733

RESULT 15
US-09-873-880-29/C
; Sequence 29, Application US/09873880
; Patent No. US20020123118A1
; GENERAL INFORMATION:
; APPLICANT: Seralit, Vincent
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
; FILE REFERENCE: B81192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Glycine max
US-09-873-880-29

Query Match 75.2%; Score 15.8; DB 10; Length 1362;
Best Local Similarity 89.5%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AGGGGATTGGAGATTGTG 20
|||
Db 360 AGGGGATTGGAGATTGCG 342

Search completed: July 8, 2003, 06:04:29
Job time : 25.6338 secs

Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21
Db 2829 TAGGGATTGGATTGATGAGA 2849

RESULT 11

US-10-172-086-11
; Sequence 11, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; TITLE REFERENCE: of prostate tumors
; CURRENT APPLICATION NUMBER: US/10/172,086
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 11
; LENGTH: 5310
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-11

Query Match

Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21
Db 3989 TAGGGGTGTGGAGATTATGA 4009

RESULT 12

US-09-873-880-7/c
; Sequence 7, Application US/09873880
; Patent No. US20020123118A1
; GENERAL INFORMATION:
; APPLICANT: Sewalt, Vincent
; APPLICANT: Faico, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
; FILE REFERENCE: B81192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (382)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (434)
; OTHER INFORMATION: n=A, C, G, or T
US-09-873-880-7

Query Match

Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTG 20

Db 353 AGGGATTGGAGATTGCG 335

RESULT 13

US-10-001-879-38/c
; Sequence 38, Application US/10001879
; Patent No. US20020127237A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Caferkey, Robert
; APPLICANT: Ali, Shujath
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes a
; FILE REFERENCE: DEX-0281
; CURRENT APPLICATION NUMBER: US/10/001,879
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,188
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 201
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 920
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-879-38

Query Match

Best Local Similarity 89.5%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 GGGGATTGGAGATTGTGA 21
Db 692 GGGGATGAGAGATTGTGA 674

RESULT 14

US-09-738-626-2066
; Sequence 2066, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIYO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2066
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2066

Query Match

75.2%; Score 15.8; DB 9; Length 1149;

```

; APPLICANT: Foy, Theresa M.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121, 479C1
; CURRENT APPLICATION NUMBER: US/09/757,417
; CURRENT FILING DATE: 2001-01-08
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-09-757-417-23

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 422;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
1 TAGGGGATTGGAGATTGTGA 21
|||||

DB
280 TCGGGGTTTGGAGATGTGA 260

RESULT 7
US-09-918-995-33366
; Sequence 33366, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33366
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-33366

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 9; Length 430;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
1 TAGGGGATTGGAGATTGTGA 21
|||||

DB
347 TAGGAGATTGGAAATTCTGA 367

RESULT 8
US-09-918-995-34536
; Sequence 34536, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34536
; LENGTH: 475
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(475)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-34536

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 9; Length 475;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
1 TAGGGGATTGGAGATTGTGA 21
|||||

DB
442 TAGGGGAAATGGAGACTGTGA 462

RESULT 9
US-09-938-842A-4643/C
; Sequence 4643, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCSRP1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4643
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4643

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 9; Length 2000;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
1 TAGGGGATTGGAGATTGTGA 21
|||||

DB
223 TAGGAGATTGGAGATTGTGA 203

RESULT 10
US-10-172-086-42
; Sequence 42, Application US/10172086
; Publication No. US20030113750A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the differentiation
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/172,086
; CURRENT FILING DATE: 2002-06-13
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 42
; LENGTH: 5118
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-172-086-42

Query Match
77.1%; Score 16.2; DB 9; Length 5118;

```

```
; CURRENT FILING DATE: 2001-10-09
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 11
; LENGTH: 21000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-975-123-11

Query Match          80.0%; Score 16.8; DB 9; Length 21000;
Best Local Similarity 90.0%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AGGGGATTGGAGATTGTA 21
        |||||
Db       1787 AGGGGATTGGAGATTGTA 1768

RESULT 3
US-09-815-242-7388
; Sequence 7388, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7388
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; NAME/KEY: CDS
; LOCATION: (1)...(981)
US-09-815-242-7388

Query Match          78.1%; Score 16.4; DB 10; Length 981;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AGGGGATTGGAGATTGT 19
        |||||
Db       810 AGGGGATTGGAGATTGT 827

RESULT 4
US-10-102-524-97
; Sequence 97, Application US/10102524
```

```
; Publication No. US20030109434A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Gaiger, Alexander
; APPLICANT: Gordon, Brian
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.572
; CURRENT APPLICATION NUMBER: US/10/102,524
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 1863
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 186
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; LOCATION: 15, 51
; OTHER INFORMATION: n = A,T,C or G
US-10-102-524-97

Query Match          77.1%; Score 16.2; DB 9; Length 186;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        |||||
Db       114 TTGGGATTGGGATTGTA 134

RESULT 5
US-10-042-945-23/C
; Sequence 23, Application US/10042945
; Publication No. US20030045468A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Steven P.
; APPLICANT: Foy, Teresa M.
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Wang, Aljun
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Sutherland, R. Alec
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY,
; FILE REFERENCE: 210121.479C3
; CURRENT APPLICATION NUMBER: US/10/042,945
; CURRENT FILING DATE: 2002-01-08
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 422
; TYPE: DNA
; ORGANISM: Oryctolagus cuniculus
US-10-042-945-23

Query Match          77.1%; Score 16.2; DB 9; Length 422;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGTA 21
        |||||
Db       280 TCGGGATTGGAGATTGTA 260

RESULT 6
US-09-757-417-23/C
; Sequence 23, Application US/09757417
; Patent No. US20020082216A1
; GENERAL INFORMATION:
; APPLICANT: Fanger, Gary R.
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 10.6338 Seconds
(without alignments)
3082.690 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21

Sequence: 1 tagggagattgagattgtga 21

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database :

Published Applications.NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCIT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCITUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	436	10	US-09-878-574-4901
2	16.8	80.0	21000	9	US-09-975-123-11
3	16.4	78.1	981	10	US-09-815-242-7388
4	16.2	77.1	186	9	US-10-102-824-97
5	16.2	77.1	422	9	US-10-042-945-23
6	16.2	77.1	422	10	US-09-757-417-23
7	16.2	77.1	475	9	US-09-918-995-3336
8	16.2	77.1	475	9	US-09-918-995-3336
9	16.2	77.1	2000	9	US-09-838-842A-4643
10	16.2	77.1	5118	9	US-10-172-086-42
11	16.2	77.1	5310	9	US-10-172-086-41
12	15.8	75.2	495	10	US-09-873-880-7
13	15.8	75.2	920	12	US-10-001-879-38
14	15.8	75.2	1149	9	US-09-738-626-2066
15	15.8	75.2	1362	10	US-09-873-880-29
16	15.8	75.2	81826	9	US-10-175-523-197
17	15.8	75.2	3509400	9	US-09-738-626-1
18	15.4	73.3	877	9	US-09-770-445-539
19	15.4	73.3	7770	10	US-09-991-630-23

20	15.4	73.3	684973	10	US-09-263-959-1	Sequence 1, Appl
21	15.2	72.4	211	10	US-09-867-701-7906	Sequence 7906, Ap
22	15.2	72.4	229	10	US-09-867-701-7696	Sequence 7696, Ap
23	15.2	72.4	257	9	US-10-040-739-122	Sequence 122, App
24	15.2	72.4	336	9	US-09-796-692-2706	Sequence 2706, Ap
25	15.2	72.4	336	9	US-10-040-662-2706	Sequence 2706, Ap
26	15.2	72.4	348	9	US-09-736-457-1382	Sequence 1382, Ap
27	15.2	72.4	348	9	US-09-902-941-1382	Sequence 1382, Ap
28	15.2	72.4	348	9	US-09-849-626-1382	Sequence 1382, Ap
29	15.2	72.4	348	9	US-10-017-754-1382	Sequence 1382, Ap
30	15.2	72.4	372	10	US-09-878-574-1079	Sequence 1079, Ap
31	15.2	72.4	382	10	US-09-878-574-2535	Sequence 2535, Ap
32	15.2	72.4	425	10	US-09-860-352-2921	Sequence 2921, Ap
33	15.2	72.4	462	10	US-09-864-761-14436	Sequence 14436, A
34	15.2	72.4	473	10	US-09-783-590-5415	Sequence 5415, Ap
35	15.2	72.4	481	9	US-09-907-969-46	Sequence 46, Appl
36	15.2	72.4	481	10	US-09-884-441-46	Sequence 46, Appl
37	15.2	72.4	486	10	US-09-864-761-5386	Sequence 5386, Ap
38	15.2	72.4	510	9	US-09-918-995-2254	Sequence 2254, Ap
39	15.2	72.4	518	9	US-09-770-961-67	Sequence 67, Appl
40	15.2	72.4	528	8	US-08-945-038-9	Sequence 9, Appl
41	15.2	72.4	580	10	US-09-864-761-21927	Sequence 21927, A
42	15.2	72.4	871	9	US-09-989-442-45	Sequence 45, Appl
43	15.2	72.4	1056	9	US-09-938-842A-2129	Sequence 2129, Ap
44	15.2	72.4	1080	9	US-09-988-442-80	Sequence 80, Appl
45	15.2	72.4	1080	9	US-10-074-045-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-878-574-4901/c
; Sequence 4901, Application US/09878574
; Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Brium, Joseph R.
APPLICANT: La Rosa, Thomas J.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(15401)B
CURRENT APPLICATION NUMBER: US/09/878, 574
CURRENT FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 09/533, 535
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 4901
LENGTH: 436
TYPE: DNA
ORGANISM: Glycine max
OTHER INFORMATION: Clone ID: LIB3028-054-01-B1-B3
US-09-878-574-4901

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 10; Length 436;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21
DB 170 TTGGGATTGGAGATTGTGA 150

RESULT 2
US-09-975-123-11/c
; Sequence 11, Application US/09975123
; Publication No. US20030087857A1
GENERAL INFORMATION:
APPLICANT: Susan M. Freier
TITLE OF INVENTION: ANTISENSE MODULATION OF INSULIN-LIKE GROWTH FACTOR BINDING PRO
FILE REFERENCE: RTS-0253
CURRENT APPLICATION NUMBER: US/09/975,123

Fri Jul 11 09:24:03 2003

us-09-477-082-31.rni

Page 8

Job time : 18.0371 secs

QY 1 TAGGGATTGGAGATTGTG 20
 ||||| ||||| ||||| |||||
 Db 1078 TAGGCATTGTAGATGTG 1059

RESULT 14
 US-09-173-891-5/c
 ; Sequence 5, Application US/09173891
 ; Patent No. 607937
 ; GENERAL INFORMATION:
 ; APPLICANT: Payne, Jewel
 ; APPLICANT: Kennedy, M. Keith
 ; APPLICANT: Randall, John Brooks
 ; APPLICANT: Meier, Henry
 ; APPLICANT: Wick, Heidi Jane
 ; APPLICANT: Foncecrada, Luis
 ; APPLICANT: Schmeff, H. Ernest
 ; APPLICANT: Schwab, George E.
 ; APPLICANT: Fu, Jenny
 ; TITLE OF INVENTION: No. 6077937e1 Bacillus thuringiensis Toxins Active
 ; TITLE OF INVENTION: Against Hymenopteran Pests
 ; NUMBER OF SEQUENCES: 51
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: David R. Saliwanichik
 ; STREET: 2421 N.W. 41st Street, Suite A-1
 ; CITY: Gainesville
 ; STATE: FL
 ; COUNTRY: USA
 ; ZIP: 32606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/173,891
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/158,232
 ; FILING DATE:
 ; APPLICATION NUMBER: US 07/887,980
 ; FILING DATE: 22-MAY-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/797,645
 ; FILING DATE: 25-NOV-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/703,977
 ; FILING DATE: 22-MAY-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Saliwanichik, David R.
 ; REGISTRATION NUMBER: 31,794
 ; REFERENCE/DOCKET NUMBER: M/SCU104.C1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 904-375-8100
 ; TELEFAX: 904-372-5800
 ; INFORMATION FOR SEQ ID NO: 5:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3771 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHEICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Bacillus thuringiensis
 ; INDIVIDUAL ISOLATE: 33F2
 ; IMMEDIATE SOURCE:
 ; CLONE: E. coli NM522(pMTC316) B-18785
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; LOCATION: 4..24

OTHER INFORMATION: /function="oligonucleotide"
 OTHER INFORMATION: hybridization probe"
 OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"
 OTHER INFORMATION: /standard_name="probe a"
 OTHER INFORMATION: /note="Probe A"
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; LOCATION: 13..33
 ; OTHER INFORMATION: /function="oligonucleotide"
 ; OTHER INFORMATION: hybridization probe"
 ; OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"
 ; OTHER INFORMATION: /standard_name="Probe B"
 ; OTHER INFORMATION: /label="probe-b"
 ; OTHER INFORMATION: /note="probe b"
 US-09-173-891-5

Query Match 72.4% Score 15.2; DB 3; Length 3771;
 Best Local Similarity 85.0% Pred. No. 2.2e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20
 ||||| ||||| ||||| |||||
 Db 1078 TAGGCATTGTAGATGTG 1059

RESULT 15
 US-09-076-137-5/c
 ; Sequence 5, Application US/09076137B
 ; Patent No. 6166195
 ; GENERAL INFORMATION:
 ; APPLICANT: Schmeff, Harry E.
 ; APPLICANT: Schwab, George E.
 ; APPLICANT: Payne, Jewel M.
 ; APPLICANT: Narva, Kenneth E.
 ; APPLICANT: Foncecrada, Luis
 ; TITLE OF INVENTION: No. 6166195e1 Nematode-Active Toxins and Genes Which Code
 ; TITLE OF INVENTION: Therefor
 ; FILE REFERENCE: MA-20CCCD2
 ; CURRENT APPLICATION NUMBER: US/09/076,137B
 ; EARLIER FILING DATE: 1998-05-12
 ; EARLIER APPLICATION NUMBER: 08/316,301
 ; EARLIER FILING DATE: 1994-09-30
 ; NUMBER OF SEQ ID NOS: 42
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 3771
 ; TYPE: DNA
 ; ORGANISM: Bacillus thuringiensis
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; LOCATION: (4)..(24)
 ; OTHER INFORMATION: /function="oligonucleotide hybridization probe"
 ; OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"
 ; OTHER INFORMATION: /standard_name="probe a" /note="Probe A"
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; LOCATION: (13)..(33)
 ; OTHER INFORMATION: /function="oligonucleotide hybridization probe"
 ; OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"
 ; OTHER INFORMATION: /standard_name="Probe B" /label="probe-b" /note-
 ; OTHER INFORMATION: "probe b"
 US-09-076-137-5

Query Match 72.4% Score 15.2; DB 4; Length 3771;
 Best Local Similarity 85.0% Pred. No. 2.2e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20
 ||||| ||||| ||||| |||||
 Db 1078 TAGGCATTGTAGATGTG 1059

Search completed: July 8, 2003, 05:56:57

FILING DATE: 12-AUG-1987
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/830,050
FILING DATE: 31-JAN-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lloyd, Jeff
REGISTRATION NUMBER: 35,589
REFERENCE/DOCKET NUMBER: MA20CCDD1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 5 (PS33F2):
SEQUENCE CHARACTERISTICS:
LENGTH: 3771 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
IMMEDIATE SOURCE: 33f2
CLONE: E. coli NM522(pMYC 2316) B-18785
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4..24
OTHER INFORMATION: /function= oligonucleotide
OTHER INFORMATION: /hybridization probe
OTHER INFORMATION: /product= GCA/T ACA/T TTA AAT GAA GTA/T TAT
OTHER INFORMATION: /standard_name= probe a
OTHER INFORMATION: /note= probe a
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13..33
OTHER INFORMATION: /function= oligonucleotide
OTHER INFORMATION: /hybridization probe
OTHER INFORMATION: /product= AAT GAA GTA/T TAT CCA/T GTA/T AAT
OTHER INFORMATION: /standard_name= Probe B
OTHER INFORMATION: /label= probe-b
OTHER INFORMATION: /note= probe b
US-08-316-301a-5
Query Match 72.4%; Score 15.2; DB 1; Length 3771;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TAGGGGATTGGAGATTGTG 20
|||||
Db 1078 TAGGGGATTGTGAAGATG 1059
|||||
RESULT 13
US-08-611-928-5/C
Sequence 5, Application US/08611928
Patent No. 5824/92
GENERAL INFORMATION:
APPLICANT: Payne, Jewel
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Uick, Heidi Jane
APPLICANT: Forcetrada, Luis
APPLICANT: Schnepf, H. Ernest
APPLICANT: Schwab, George E.
APPLICANT: Fu, Jenny
TITLE OF INVENTION: No. 5824/92a1 Bacillus thuringiensis Toxins Active
TITLE OF INVENTION: Against Hymenopteran Pests
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Salwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,928
FILING DATE: 06-MAR-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,232
FILING DATE: 24-NOV-1993
APPLICATION NUMBER: US 07/887,980
FILING DATE: 22-MAY-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/797,645
FILING DATE: 25-NOV-1991
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/703,977
FILING DATE: 22-MAY-1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Salwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ104.C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3771 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: 33f2
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC2316) B-18785
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4..24
OTHER INFORMATION: /function= oligonucleotide
OTHER INFORMATION: /hybridization probe
OTHER INFORMATION: /product= GCA/T ACA/T TTA AAT GAA GTA/T TAT
OTHER INFORMATION: /standard_name= "probe a"
OTHER INFORMATION: /note= "probe a"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13..33
OTHER INFORMATION: /function= oligonucleotide
OTHER INFORMATION: /hybridization probe
OTHER INFORMATION: /product= AAT GAA GTA/T TAT CCA/T GTA/T AAT
OTHER INFORMATION: /standard_name= "Probe B"
OTHER INFORMATION: /label= probe-b
OTHER INFORMATION: /note= "probe b"
US-08-611-928-5
Query Match 72.4%; Score 15.2; DB 1; Length 3771;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;


```
LOCATION: 13..33 /function="oligonucleotide"
OTHER INFORMATION: hybridization probe"
OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"
OTHER INFORMATION: /standard_name="Probe B"
OTHER INFORMATION: /label="probe-b"
OTHER INFORMATION: /note="probe b"
US-08-158-232-5

Query Match
Best Local Similarity 72.4%; Score 15.2; DB 1; Length 3771;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTG 20
Db 1078 TAGGGGATTGTGAGATTGTG 1059

RESULT 11
US-08-304-626-5/c
Sequence 5, Application US/08304626
Patent No. 5616495
GENERAL INFORMATION:
APPLICANT: Payne, Jewel M.
APPLICANT: Kennedy, M. Keith
APPLICANT: Randall, John Brooks
APPLICANT: Meier, Henry
APPLICANT: Ulick, Heidi Jane
APPLICANT: Foncerrada, Luis
APPLICANT: Schrepf, Harry E.
APPLICANT: Schwab, George E.
TITLE OF INVENTION: No. 5616495el Bacillus thuringiensis Isolates
TITLE OF INVENTION: Active Against Hymenopteran Pests and Genes Encoding
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Salivanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/304,626
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/887,980
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Salivanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/SCJ 104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3771 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOCHERICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
INDIVIDUAL ISOLATE: 33F2
```

```
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMKC2316) B-18785
FEATURE:
NAME/KEY: misc feature
LOCATION: 4..24
OTHER INFORMATION: /function="oligonucleotide"
OTHER INFORMATION: hybridization probe"
OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"
OTHER INFORMATION: /standard_name="probe a"
OTHER INFORMATION: /note="probe A"
FEATURE:
NAME/KEY: misc feature
LOCATION: 13..33
OTHER INFORMATION: /function="oligonucleotide"
OTHER INFORMATION: hybridization probe"
OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"
OTHER INFORMATION: /standard_name="Probe B"
OTHER INFORMATION: /label="probe-b"
OTHER INFORMATION: /note="probe b"
US-08-304-626-5

Query Match
Best Local Similarity 72.4%; Score 15.2; DB 1; Length 3771;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTG 20
Db 1078 TAGGGGATTGTGAGATTGTG 1059

RESULT 12
US-08-316-301A-5/c
Sequence 5, Application US/08316301A
Patent No. 5753492
GENERAL INFORMATION:
APPLICANT: Schrepf, Harry E.
APPLICANT: Schwab, George E.
APPLICANT: Payne, Jewel M.
APPLICANT: Narva, Kenneth E.
APPLICANT: Foncerrada, Luis
TITLE OF INVENTION: No. 5753492el Nematode-Active Toxins and Genes
TITLE OF INVENTION: Which Code Therefor
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSER: Salivanchik & Salivanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,301A
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/871,510
FILING DATE: 23-APR-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/693,018
FILING DATE: 03-MAY-1991
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/565,544
FILING DATE: 10-AUG-1990
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/084,653
```

US-08-049-783-1/C
; Sequence 1, Application US/08049783
; Patent No. 5439881
; GENERAL INFORMATION:
; APPLICANT: Narva, Kenneth E
; APPLICANT: Schwab, George E
; APPLICANT: Payne, Jewel M
; TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active
; TITLE OF INVENTION: Toxins Cloned from Bacillus thuringiensis Isolates
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 1 (PS33F2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: 33f2a
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 4..24
; OTHER INFORMATION: /function="oligonucleotide
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"
; OTHER INFORMATION: /standard_name="probe a"
; OTHER INFORMATION: /note="probe A"
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 13..33
; OTHER INFORMATION: /function="oligonucleotide
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product="AAT GAA GTA/T TAT CCA/T GTA/T AAT"
; OTHER INFORMATION: /standard_name="probe B"
; OTHER INFORMATION: /label="probe-b"
; OTHER INFORMATION: /note="probe b"
US-08-049-783-1
Query Match 72.4%; Score 15.2; DB 1; Length 3771;
Best Local Similarity 85.0%; Pred. No. 2.2e-02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
CY 1 TAGGGATTGGAGATTG 20
DB 1078 TAGGGATTGGAGATTG 1059

RESULT 10
US-08-158-232-5/C
; Sequence 5, Application US/08158232
; Patent No. 5596071
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Kennedy, M. Keith
; APPLICANT: Randall, John Brooks
; APPLICANT: Meier, Henry
; APPLICANT: Vick, Heidi Jane
; APPLICANT: Fonceriada, Luis
; APPLICANT: Schepf, H. Ernest
; APPLICANT: Schwab, George E.
; APPLICANT: Fu, Jenny
; TITLE OF INVENTION: No. 5596071el Bacillus thuringiensis Toxins Active
; TITLE OF INVENTION: Against Hymenopteran Pests
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,232
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/887,980
; FILING DATE: 22-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/797,645
; FILING DATE: 25-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/703,977
; FILING DATE: 22-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/SCI104.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC316) B-18785
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 4..24
; OTHER INFORMATION: /function="oligonucleotide
; OTHER INFORMATION: hybridization probe"
; OTHER INFORMATION: /product="GCA/T ACA/T TTA AAT GAA GTA/T TAT"
; OTHER INFORMATION: /standard_name="probe a"
; OTHER INFORMATION: /note="probe A"
; FEATURE:
; NAME/KEY: misc-feature

RESULT 6
US-09-262-773-5
; Sequence 5, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 3264
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (98)..(496)
US-09-262-773-5

Query Match 72.4%; Score 15.2; DB 4; Length 3264;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGCA 21
|||||
Db 2993 AGGGATTGGAACTGCA 3012

RESULT 7
US-09-262-773-1
; Sequence 1, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3268
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (98)..(2041)
US-09-262-773-1

Query Match 72.4%; Score 15.2; DB 4; Length 3268;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGCA 21
|||||
Db 2997 AGGGATTGGAACTGCA 3016

RESULT 8
US-07-876-280-5/c
; Sequence 5, Application US/07876280
; Patent No. 5262158

GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; APPLICANT: Cannon, Raymond J.C.
; APPLICANT: Bagley, Angela L.
; TITLE OF INVENTION: No. 5262158el Bacillus thuringiensis Isolates for
; TITLE OF INVENTION: Controlling Acarides
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/876,280
; FILING DATE: 19920430
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: M/S 104
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3771 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: 33f2
; IMMEDIATE SOURCE:
; CLONE: E. coli NM522(pMYC 2316) B-18785
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 4..24
; OTHER INFORMATION: /function= "oligonucleotide"
; OTHER INFORMATION: /hybridization probe"
; OTHER INFORMATION: /product= "GCA/T ACA/T TTA AAT GAA GTA/T TAT"
; OTHER INFORMATION: /standard_name= "probe a"
; OTHER INFORMATION: /note= "probe a"
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13..33
; OTHER INFORMATION: /function= "oligonucleotide"
; OTHER INFORMATION: /hybridization probe"
; OTHER INFORMATION: /product= "AAT GAA GTA/T TAT CCA/T GTA/T AAT"
; OTHER INFORMATION: /standard_name= "probe b"
; OTHER INFORMATION: /label= "probe-b"
; OTHER INFORMATION: /note= "probe b"
US-07-876-280-5

Query Match 72.4%; Score 15.2; DB 1; Length 3771;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGCGATTGGAGATTGTG 20
|||||
Db 1078 TAGGCGATTGTAGATTGTG 1059

RESULT 9

APPLICANT: Algate, Paul A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.462C2
CURRENT APPLICATION NUMBER: US/09/404,879A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 393
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 46
LENGTH: 481
TYPE: DNA
ORGANISM: Homo sapien
US-09-404-879A-46

Query Match 72.4%; Score 15.2; DB 4; Length 481;
Best Local Similarity 85.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20
Db 117 TAGGCTATTGGAGATTGTG 98

RESULT 3
US-08-993-228-20
Sequence 20, Application US/08993228
Patent No. 5976838
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallee, Edward R.
APPLICANT: Racine, Lisa A.
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 33
TITLE OF INVENTION: ENCODING THEM
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,228
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 3113 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-993-228-20

Query Match 72.4%; Score 15.2; DB 2; Length 3113;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGCA 21
Db 407 AGGGAAATTGGAGCTGTGCA 426

RESULT 4
US-09-262-773-7
Sequence 7, Application US/09262773
Patent No. 6225451
GENERAL INFORMATION:
APPLICANT: Ballinger, Dennis G.
APPLICANT: Ding, Wei
APPLICANT: Wagner, Susanne
APPLICANT: Hess, Mark A.
TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
FILE REFERENCE: Myriad 3
CURRENT APPLICATION NUMBER: US/09/262,773
CURRENT FILING DATE: 1999-03-04
NUMBER OF SEQ ID NOS: 210
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 3240
TYPE: DNA
ORGANISM: human
FEATURE:
NAME/KEY: CDS
LOCATION: (98)..(523)
US-09-262-773-7

Query Match 72.4%; Score 15.2; DB 4; Length 3240;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGCA 21
Db 2969 AGGGATTGGAACTGTCA 2988

RESULT 5
US-09-262-773-3
Sequence 3, Application US/09262773
Patent No. 6225451
GENERAL INFORMATION:
APPLICANT: Ballinger, Dennis G.
APPLICANT: Ding, Wei
APPLICANT: Wagner, Susanne
APPLICANT: Hess, Mark A.
TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
FILE REFERENCE: Myriad 3
CURRENT APPLICATION NUMBER: US/09/262,773
CURRENT FILING DATE: 1999-03-04
NUMBER OF SEQ ID NOS: 210
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 3244
TYPE: DNA
ORGANISM: human
FEATURE:
NAME/KEY: CDS
LOCATION: (98)..(2017)
US-09-262-773-3

Query Match 72.4%; Score 15.2; DB 4; Length 3244;
Best Local Similarity 85.0%; Pred. No. 2.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGCA 21
Db 2973 AGGGATTGGAACTGTCA 2992

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 5.03708 Seconds

(without alignments)
1278.561 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21

Sequence: 1 taggggattggagattgtga 21

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA.*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	16.8	80.0	168575	4 US-09-426-290-1
2	15.2	72.4	481	4 US-09-404-879A-46
3	15.2	72.4	3113	2 US-08-993-228-20
4	15.2	72.4	3240	4 US-09-262-773-7
5	15.2	72.4	3244	4 US-09-262-773-3
6	15.2	72.4	3264	4 US-09-262-773-5
7	15.2	72.4	3268	4 US-09-262-773-1
8	15.2	72.4	3771	1 US-07-876-280-5
9	15.2	72.4	3771	1 US-08-049-783-1
10	15.2	72.4	3771	1 US-08-158-232-5
11	15.2	72.4	3771	1 US-08-304-626-5
12	15.2	72.4	3771	1 US-08-316-301A-5
13	15.2	72.4	3771	1 US-08-611-928-5
14	15.2	72.4	3771	3 US-09-173-891-5
15	15.2	72.4	3771	4 US-09-076-137-5
16	15.2	72.4	3771	5 PCT-US92-03624-5
17	15.2	72.4	5849	4 US-09-134-246-6
18	15.2	72.4	20137	4 US-09-262-773-206
19	15.2	72.4	20138	4 US-09-262-773-9
20	15.2	72.4	23071	4 US-09-262-773-210
21	15.2	72.4	15231	3 US-09-128-155-16
22	15.2	72.4	176373	3 US-09-128-155-17
23	14.8	70.5	22	4 US-09-262-773-187
24	14.8	70.5	543	4 US-09-221-017B-681
25	14.8	70.5	1340	4 US-09-000-062-2
26	14.8	70.5	1340	4 US-09-000-062-4
27	14.8	70.5	1713	4 US-09-000-062-1

28	14.8	70.5	1978	6 5188642-3	Patent No. 5188642
29	14.8	70.5	2643	4 US-09-399-913-56	Sequence 56, Appl
30	14.8	70.5	21040	4 US-08-961-527-55	Sequence 55, Appl
31	14.6	69.5	407	4 US-09-605-785-506	Sequence 506, Appl
32	14.6	69.5	407	4 US-09-439-313-506	Sequence 506, Appl
33	14.6	69.5	411	4 US-09-605-785-508	Sequence 508, Appl
34	14.6	69.5	411	4 US-09-439-313-508	Sequence 508, Appl
35	14.6	69.5	422	4 US-09-605-785-507	Sequence 507, Appl
36	14.6	69.5	422	4 US-09-439-313-507	Sequence 507, Appl
37	14.6	69.5	768	1 US-08-592-126-116	Sequence 116, Appl
38	14.6	69.5	1693	4 US-09-149-476-224	Sequence 224, Appl
39	14.6	69.5	1753	4 US-09-149-476-56	Sequence 56, Appl
40	14.6	69.5	2380	3 US-08-700-651-4	Sequence 4, Appl
41	14.6	69.5	2887	4 US-08-383-502-14	Sequence 14, Appl
42	14.6	69.5	2887	5 PCT-US96-10521-14	Sequence 14, Appl
43	14.6	69.5	28473	4 US-08-961-527-83	Sequence 83, Appl
44	14.4	68.6	256	2 US-08-365-486A-22	Sequence 22, Appl
45	14.4	68.6	256	4 US-08-880-342-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-09-426-290-1
Sequence 1, Application US/09426290
Patent No. 6410712
GENERAL INFORMATION:
APPLICANT: Berglind Ran Olafsdottir
TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
FILE REFERENCE: 2345.2001-000
CURRENT APPLICATION NUMBER: US/09/426.290
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 168575
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101966)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (128910)...(129139)
US-09-426-290-1
Query Match 80.0%; Score 16.8; DB 4; Length 168575;
Best Local Similarity 90.0%; Pred. No. 57;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Db 29456 AGGGATTGGAGATTGTGA 21
OY 2 AGGGATTGGAGATTGTGA 21
RESULT 2
US-09-404-879A-46/c
Sequence 46, Application US/09404879A
Patent No. 6468546
GENERAL INFORMATION:
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3974043"
/tissue="Scars_NMEB_branchial_arch"
/dev_stage="embryo, 10.5 dpc"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pUT3D-Pac (Pharmacia) with a modified
polylinker; site_1: NotI; site_2: EcoRI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGACGCGGCGCGCATGTTTGTGTTTGTGTTTGTGTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pUT3D vector. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT      114 a      130 c      191 g      110 t      1 others
ORIGIN

```

```

Query Match      82.9%; Score 17.4; DB 12; Length 546;
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      2 AGGGGATTGGAGATTGTG 20
        |||||
Db      477 AGGGGATTGGAGATTGTG 495

```

Search completed: July 8, 2003, 05:51:38
 Job time : 174.34 secs

RESULT 13
BE471807 488 bp mRNA linear EST 28-JUL-2000
LOCUS EST16660 potato stolon, Cornell University Solanum tuberosum cDNA
DEFINITION BE471807
ACCESSION BE471807
VERSION BE471807.1 GI:9562298
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 488)
van der Hoeven,R., Bezzerides,J., Bachem,C., Horvath,B., Visser,R., Holt,I.E., Liang,F., Hansen,T.S., Uiterback,T., Bowman,C.L., Doan,B., Bougri,O., Buell,C.R., Romling,C.M., Tanksley,S.D. and Baker,B.
TITLE Generation of ESTs from potato swelling stolons
JOURNAL Unpublished (1999)
COMMENT Contact: Research Genetics, Libraries Division
Tel: 1-800-711-6195
Email: cdnaresgen.com
5 prime sequence.
FEATURES
source Location/Qualifiers
1..488
/organism="Solanum tuberosum"
/cultivar="Binje"
/db_xref="taxon:4113"
/clone="GST30H15"
/clone_lib="potato stolon, Cornell University"
/tissue_type="axillary buds of stem explants, swelling stolons"
/dev_stage="1 to 3 days"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; RNA was supplied by Christian Bachem & Beatrix Horvath(Laboratory of Plant Breeding, Dept. of Plant Sciences, Wageningen University, The Netherlands). Total RNA was isolated from developing axillary buds of potato nodal stem cuttings cultured on medium for the introduction of tuber formation as described in Bachem et al. (Plant Journal 1996). Tissue samples were taken of stages corresponding to growing stolons and the early stages of tuber formation."
BASE COUNT 99 a 108 c 136 g 145 t
ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 488;
Best Local Similarity 94.7%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTG 20
|||||
Db 414 AGGGATTGGAGATTGTG 432

RESULT 14
BQ118560 535 bp mRNA linear EST 22-JUL-2002
LOCUS EST604136 mixed potato tissues, Solanum tuberosum cDNA clone STMB75
DEFINITION BQ118560
ACCESSION BQ118560
VERSION BQ118560.2 GI:21918968
KEYWORDS EST.
SOURCE potato.
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 535)
Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and

TITLE Karamycheva,S.A.
JOURNAL Generation of a set of potato cDNA clones for microarray analyses
COMMENT Unpublished (2002)
On Apr 17, 2002 this sequence version replaced gi:20170522.
Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potato@tigr.org
This clone is available through the Research Genetics, contact the Research Genetics for further information 1-800-711-6195 or cdnaresgen.com
Seq primer: T3.
FEATURES
source Location/Qualifiers
1..535
/organism="Solanum tuberosum"
/cultivar="Kennebec or Binjite"
/db_xref="taxon:4113"
/clone="STMB75"
/clone_lib="mixed potato tissues"
/tissue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: combination of untreated and phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes, tubers, or roots."
BASE COUNT 105 a 121 c 149 g 160 t
ORIGIN

Query Match 82.9%; Score 17.4; DB 14; Length 535;
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGGATTGGAGATTGTG 20
|||||
Db 414 AGGGATTGGAGATTGTG 432

RESULT 15
BQ147712 546 bp mRNA linear EST 01-FEB-2001
LOCUS mab53402.x1 Soares_NMEBA_branch1_arch Mus musculus cDNA clone
DEFINITION IMAGE:3974043 3 similar to IR:Q9V115 Q9V115 CG10286 PROTEIN.?, mRNA sequence.
ACCESSION BQ147712
VERSION BQ147712.1 GI:12651120
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 546)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Other-ESTs: mab53402.y1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA library Preparation: M. Bento Soares, Ph.D., M. Fatima Ronaldo, Ph.D.
cDNA library Arranged by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov/image/html/resources.shtml
MG1:1474075
Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyt not found
High quality sequence stop: 488.
Location/Qualifiers
1..546
source

QY 1 TAGGGGATTGGAGATTGTGA 21
 ||||| |||||
 Db 497 TAGGGGATTGGAGATTGTGA 477

RESULT 6
 BU002845 607 bp mRNA linear EST 22-AUG-2002
 LOCUS OGG33P01.YG.ab1 OG_EFGHJ lettuce serriola Lactuca sativa cDNA clone

DEFINITION OGG33P01.YG.ab1 OG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
 OGG33P01 mRNA sequence.

ACCESSION BU002845
 VERSION BU002845.1 GI:22437240
 KEYWORDS EST.
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

REFERENCE 1 (bases 1 to 607)
 Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmundo Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@catc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA.Contig1855, see http://cgpdb.ucdavis.edu/
 for details
 Plate: OGG33 row: D column: 01.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1. 607
 /organism="Lactuca sativa"
 /cultivar="L.serriola"
 /db_xref="taxon:4236"
 /clone="OGG33P01"
 /clone_lib="OG_EFGHJ lettuce serriola"
 /lab_host="E.coli"
 /note="Vector: pBRCDNA51AB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_LIB-OG_EFGHJ lettuce serriola
 TAG_TISSUE=germinating seeds
 TAG_SEQ=TCGTGCGGG"

BASE COUNT 145 a 136 c 140 g 186 t
 ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 607;
 Best Local Similarity 90.5%; Pred. No. 9.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21
 ||||| |||||
 Db 56 TTGGCGATTGGAGATTGTGA 76

RESULT 7
 BO996610 630 bp mRNA linear EST 22-AUG-2002
 LOCUS BO996610

DEFINITION OGG33P01.YG.ab1 OG_EFGHJ lettuce serriola Lactuca sativa cDNA clone
 OGG33P01 mRNA sequence.
 ACCESSION BO996610
 VERSION BO996610.1 GI:22431006
 KEYWORDS EST.
 SOURCE Lactuca sativa.
 ORGANISM Lactuca sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids II; Asterales; Asteraceae; Lactuceae;
 Lactuca.

REFERENCE 1 (bases 1 to 630)
 Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L.,
 Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Asmundo Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@catc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig OG_CA.Contig1855, see http://cgpdb.ucdavis.edu/
 for details
 Plate: OGG33 row: F column: 01.

TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1. 630
 /organism="Lactuca sativa"
 /cultivar="L.serriola"
 /db_xref="taxon:4236"
 /clone="OGG33P01"
 /clone_lib="OG_EFGHJ lettuce serriola"
 /lab_host="E.coli"
 /note="Vector: pBRCDNA51AB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_LIB-OG_EFGHJ lettuce serriola
 TAG_TISSUE=germinating seeds
 TAG_SEQ=TCGTGCGGG"

BASE COUNT 149 a 140 c 148 g 193 t
 ORIGIN

Query Match 84.8%; Score 17.8; DB 14; Length 630;
 Best Local Similarity 90.5%; Pred. No. 9.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTGA 21
 ||||| |||||
 Db 56 TTGGCGATTGGAGATTGTGA 76

RESULT 8
 BH541998 665 bp DNA linear GSS 14-DEC-2001
 LOCUS BH541998/c
 DEFINITION BH541998 BOGG Brassica oleracea genomic clone BOGG888, DNA
 sequence.
 ACCESSION BH541998
 VERSION BH541998.1 GI:17793048
 KEYWORDS GSS.
 SOURCE Brassica oleracea.
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

High quality sequence stop: 527.

FEATURES
Location/Qualifiers

1..527
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="MPMGp637_13F2:MPMGp637F0213"
/clone_id="zebrafish shield stage whole embryo CDNA library MPMGP637"
/tissue_type="whole embryo"
/dev_stage="shield stage, 6 hrs post-fertilisation"
/lab_host="E.coli, XLI blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo-dT-NotI primed, SalI adaptors, directionally cloned, library normalised by oligonucleotide fingerprinting"

BASE COUNT
155 a 120 c 143 g 108 t 1 othersBASE COUNT
209 a 150 c 163 g 184 tQuery Match
Best Local Similarity 87.6%; Score 18.4; DB 13; Length 527;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;Query Match
Best Local Similarity 87.6%; Score 18.4; DB 17; Length 706;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;QY
2 AGGGGATTGGAGATTGTGA 21
|||||QY
2 AGGGGATTGGAGATTGTGA 21
|||||RESULT 2
A2507082 706 bp DNA linear GSS 05-OCT-2000
LOCUS
DEFINITION
1M0348K01R Mouse 10kb plasmid U08C1M library Mus musculus genomic
clone U08C1M0348K01 R, DNA sequence.RESULT 3
B0899695 939 bp mRNA linear EST 16-AUG-2002
LOCUS
DEFINITION
AGENCOURT_8750911 NIH_MGC_130 Mus musculus CDNA clone IMAGE:6336113
5', mRNA sequence.ACCESSION
A2507082
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.ACCESSION
B0899695
VERSION
KEYWORDS
SOURCE
ORGANISM
house mouse.REFERENCE
AUTHORS
1 (bases 1 to 706)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.REFERENCE
AUTHORS
1 (bases 1 to 939)
NIH-MGC <http://imgc.ncl.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999).TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: rdunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0348 row: K column: 01
Seq primer: CACACAGAAACACACTATGAC
Class: plasmid ends
High quality sequence stop: 706.TITLE
CDNA Library Preparation: Resgen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL13799 row: J column: 18
High quality sequence stop: 634.

FEATURES

FEATURES

Source

Source

1..706
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U08C1M0348K01"
/clone_id="Mouse 10kb plasmid U08C1M library"
/sex="Male"
/lab_host="E. coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: pMDA2nv. Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(<http://www.jax.org/resources/documents/dnares/>). The DNA
was hydrodynamically sheared by repeated passage through a

0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMDA2 (G1473211415B1AF19072.1) a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT
250 a 197 c 281 g 210 t 1 others

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 ; Search time 158.34 Seconds

(without alignments)
2147.949 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21
Sequence: 1 tagggattgagattgtga 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

EST.*
1: em_estda:*
2: em_estnu:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_liv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18.4	87.6	527	13	B1891021
2	18.4	87.6	706	17	A2507082
3	18.4	87.6	939	14	BQ898955
4	18.4	87.6	715	13	B198770
5	17.8	84.8	512	17	AQ862416
6	17.8	84.8	607	14	BQ002845

Result No.	Score	Query Match	Length DB	ID	Description
7	17.8	84.8	630	14	BQ966610
8	17.8	84.8	665	17	BH541998
9	17.8	84.8	710	14	BQ986827
10	17.8	84.8	737	14	BQ997349
11	17.8	84.8	742	14	BQ859778
12	17.8	84.8	748	14	BQ004608
13	17.4	82.9	488	10	BE471807
14	17.4	82.9	535	14	BQ118560
15	17.4	82.9	546	12	BG147712
16	17.4	82.9	554	10	AW735936
17	17.4	82.9	566	13	B1479559
18	17.4	82.9	621	12	BQ597208
19	17.4	82.9	649	13	BQ078834
20	17.4	82.9	685	17	AG140974
21	17.4	82.9	688	14	BQ027185
22	17.4	82.9	696	14	BQ045836
23	17.4	82.9	714	17	AG092683
24	17.4	82.9	737	13	BK113418
25	17.4	82.9	743	12	BQ597487
26	17.4	82.9	767	14	BQ113951
27	17.4	82.9	1101	17	CNS00KAG
28	17.4	82.9	1151	14	BK914796
29	17.4	82.9	663	13	B1271871
30	17.4	82.9	706	13	B1958820
31	17.4	82.9	1101	17	CNS00G9U
32	16.8	80.0	262	12	BF772261
33	16.8	80.0	292	9	AV166017
34	16.8	80.0	351	13	B1051652
35	16.8	80.0	353	17	AQ045614
36	16.8	80.0	367	10	AW315961
37	16.8	80.0	374	12	BE976571
38	16.8	80.0	396	9	A1507111
39	16.8	80.0	407	17	BH529081
40	16.8	80.0	407	9	A1736949
41	16.8	80.0	413	12	BE993537
42	16.8	80.0	438	17	BH656193
43	16.8	80.0	441	9	A1998088
44	16.8	80.0	443	17	AQ216417
45	16.8	80.0	444	17	B72074

ALIGNMENTS

RESULT 1
LOCUS B1891021
DEFINITION B1891021 527 bp mRNA EST 12-OCT-2001
ACCESSION Z6637-3-000276 zebrafish shield stage whole embryo cDNA library
VERSION MPMGP637 Danio rerio cDNA clone MPMGP637_13p2;MPMGP637F0213 5',
KEYWORDS mRNA sequence.
SOURCE B1891021.1 GI:16098292
ORGANISM zebrafish.
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 527)
AUTHORS Clark, M., Aanstad, P., Hennig, S., Johnson, S.L. and Lehrach, H.
TITLE EST sequencing of a zebrafish shield stage cDNA library normalised
by oligonucleotide fingerprinting
JOURNAL Unpublished (2001)
COMMENT Contact: Hennig S
Laboraty 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
5' EST sequencing of clones from a zebrafish shield stage library,
normalised from 55,000 starting clones by oligonucleotide
fingerprinting

CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 1199 BP; 335 A; 360 C; 320 G; 184 T; 0 other:

Query Match 80.0%; Score 16.8; DB 23; Length 1199;
 Best Local Similarity 90.0%; Pred. No. 2.3e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGTG 20
 | ||||| ||||| |||||
 Db 1012 TGGGGATGTGAGATTGTG 993

RESULT 15

AAA97385/c
 ID AAA97385 standard; DNA; 2325 BP.

XX
 AC AAA97385;

XX
 DT 29-JAN-2001 (first entry)

XX
 DE Pea p1a2 gene light-repressible promoter.

XX
 KW GTP-binding protein p1a2; pea; light-repressible promoter;

XX
 KW photoinhibitory; expression cassette; transgenic plant;

XX
 KW deterioration prevention; storage; ds.

XX
 OS Pisum sativum.

XX
 PN WO200055313-A1.

XX
 PD 21-SEP-2000.

XX
 PF 03-MAR-2000; 2000WO-JP01269.

XX
 PR 12-MAR-1999; 99JP-0066551.

XX
 PA (SUNR) SUNTORY LTD.

XX
 PI Sasaki Y, Nagano Y, Inaba T;

XX
 DR WPI; 2000-587526/55.

XX
 PT New DNA fragment or promoter for expressing a target gene, specifically
 PT under photoinhibitory conditions, and for transforming a plant cell or
 PT plant to improve quality and prevent deterioration during storage -

XX
 PS Claim 3; Page 34-35; 49pp; Japanese.

XX
 CC The invention relates to a light-repressible promoter (AAA97385), or
 CC active fragments thereof (AAA97383, AAA97384), from the gene encoding
 CC the pea GTP-binding protein p1a2. The invention also relates to an
 CC expression cassette containing the p1a2 promoter or its active
 CC fragments for the expression of a gene under photoinhibitory or dark
 CC conditions in a plant, and to transgenic plants, their descendants
 CC and plant tissues comprising the expression cassette. The expression
 CC cassette of the invention can be used to generate transgenic plants in
 CC which deterioration during storage in the dark is prevented. This is
 CC particularly useful for agricultural products. The present sequence
 CC represents the pea p1a2 promoter.

XX
 SQ Sequence 2325 BP; 791 A; 451 C; 317 G; 766 T; 0 other:

Query Match 80.0%; Score 16.8; DB 21; Length 2325;
 Best Local Similarity 90.0%; Pred. No. 2.4e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGGAGATTGTGA 21
 | ||||| ||||| |||||
 Db 2027 AAGGGATTGGAAATTGTGA 2008

Search completed: July 8, 2003, 03:11:16
 Job time : 21.9513 secs

Query Match	80.0%;	Score 16.8;	DB 21;	Length 898;
Best Local Similarity	90.0%;	Pred. No. 2.2e+02;		
Matches	18;	Conservative	0;	Mismatches 2;
			Indels	0;
			Gaps	0

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of


```

XX OS Homo sapiens.
XX PN WO200200928-A2.
XX PD 03-JAN-2002.
XX PF 02-JUL-2001; 2001WO-EP07537.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-110909/17.
XX PT Nucleic acid comprising fragment of chemically modified gene, useful
XX PT for diagnosis and treatment of diseases associated with abnormal
XX PT cytosine methylation.
XX PS Claim 1; SEQ ID NO 1435; 32pp + Sequence Listing; German.
XX CC The present invention provides a number of human immune system associated
XX CC genes which are modified by the methylation of cytosines. The sequences
XX CC can be used in the diagnosis and treatment of immune system disorders,
XX CC including eye diseases such as retinopathy, neovascular glaucoma and
XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX CC diseases. The present sequence is a gene of the invention.
XX SQ Sequence 13377 BP; 3915 A; 146 C; 3046 G; 6270 T; 0 other;

Query Match      84.8%; Score 17.8; DB 24; Length 13377;
Best Local Similarity 90.5%; Pred. No. 98;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGCA 21
DB 1704 TAGGGATTGGAGATTGTTA 1724

RESULT 11
ABL70316
ID ABL70316 standard; DNA; 16914 BP.
XX AC ABL70316;
XX DT 01-JUL-2002 (first entry)
XX DE Chemically treated cell signalling DNA sequence complementary to #103.
XX KW Cell signalling; cytosine methylation; cell signalling disease;
XX KW cancer; tumour; cytostatic; ds.
XX OS Unidentified.
XX PN WO200202807-A2.
XX PD 10-JAN-2002.
XX PF 29-JUN-2001; 2001WO-EP07471.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-154758/20.

```

```

XX OS Nucleic acid, useful for diagnosis and therapy of diseases associated
XX PT with cell signalling e.g. cancer, comprises chemically modified genomic
XX PT sequences of genes associated with cell signalling.
XX PS Claim 1; SEQ ID NO 206; 24pp+sequence listing; English.
XX CC The invention relates to a nucleic acid comprising a sequence of at least
XX CC 18 bases of a segment of chemically pre-treated DNA of genes associated
XX CC with cell signalling. The activity of the modified sequences of the
XX CC invention may be described as cytostatic. The object of the invention is
XX CC to provide the chemically modified DNA of genes associated with cell
XX CC signalling, as well as oligonucleotides and/or PNA-oligomers for
XX CC detecting cytosine methylations, as well as a method which is
XX CC particularly suitable for the diagnosis and/or therapy of genetic and
XX CC epigenetic parameters of genes associated with cell signalling. The
XX CC chemically modified DNA provided by the invention is useful for diagnosis
XX CC and therapy of diseases such as solid tumours and cancer. The sequences
XX CC given in records ABL70111-ABL70626 represent chemically pre-treated
XX CC genomic DNA's of genes associated with cell signalling.
XX CC Note: The sequence data for this patent is not represented in the printed
XX CC specification, but is based on sequence information supplied by the
XX CC European Patent Office.

SQ Sequence 16914 BP; 4573 A; 449 C; 4019 G; 7823 T; 50 other;

Query Match      81.0%; Score 17; DB 24; Length 16914;
Best Local Similarity 100.0%; Pred. No. 2,3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATT 17
DB 5123 TAGGGGATTGGAGATT 5139

RESULT 12
AAS61254
ID AAS61254 standard; DNA; 16914 BP.
XX AC AAS61254;
XX DT 29-JAN-2002 (first entry)
XX DE Human gene regulation-associated gene oligonucleotide #209.
XX KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
XX KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
XX KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
XX KW renal disease; Preeclampsia; cardiac allograft vascular disease;
XX KW colorectal cancer; thyroid cancer; oesophageal cancer; ds, tumour;
XX KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
XX KW nephrotoxic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX OS Homo sapiens.
XX PN WO200177375-A2.
XX PD 18-OCT-2001.
XX PF 06-APR-2001; 2001WO-EP03968.
XX PR 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPiG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-017470/02.
XX PT New nucleic acid sequences from chemically modified genes associated

```


CC tumour in which a myc gene is amplified, such as a neuroblastoma.
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
CC uterine cervical carcinoma can be diagnosed with the new method. A kit
CC for screening for a compound that induces death-receptor-mediated
CC apoptosis in cells containing an inactivated CASP8 gene is also
CC provided.

CC Sequence 21 BP; 5 A; 2 C; 9 G; 5 T; 0 other;

Query Match 84.8%; Score 17.8; DB 21; Length 21;
Best Local Similarity 90.5%; Pred. No. 57;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTGGAGATTGTGA 21
DB 1 TAGGGGATTGGAGATTGTGA 21

RESULT 6

ABL22738 ID ABL22738 standard; DNA; 2204 BP.

XX ABL22738;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 19687.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX Claim 1; SEQ ID NO 19687; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2204 BP; 613 A; 518 C; 507 G; 566 T; 0 other;

XX Query Match 84.8%; Score 17.8; DB 23; Length 2204;

XX Best Local Similarity 90.5%; Pred. No. 84;

XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTGGAGATTGTGA 21
DB 1558 TTGGGGATTGGAGATTGTGA 1578

RESULT 7
ABL12786 ID ABL12786 standard; CDNA; 2561 BP.

XX ABL12786;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 32840.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PADB; ABB68683.

XX Claim 1; SEQ ID NO 32840; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA

CC sequences (AB101840-AB16175) and the encoded proteins

CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2561 BP; 750 A; 531 C; 533 G; 747 T; 0 other;

XX Query Match 84.8%; Score 17.8; DB 23; Length 2561;

XX Best Local Similarity 90.5%; Pred. No. 85;

XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

XX OY 1 TAGGGGATTGGAGATTGTGA 21

XX DB 225 TTGGGGATTGGAGATTGTGA 245

RESULT 8

ABL13136 ID ABL13136 standard; CDNA; 3536 BP.

XX ABL13136;

XX 26-MAR-2002 (first entry)

CC the specification such as ACCPN, ADFN, or AFDI and comprising one of 350
 CC sequences (ABN79984-ABN80333) or their complements. The invention is
 CC useful for the diagnosis or therapy of diseases associated with
 CC development genes, in particular disease related to homeobox containing
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Curraño syndrome, diseases related with the
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC Oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all CpG dinucleotides within the 350 sequences or
 CC (II) and their complementary sequences, as primer oligonucleotides for
 CC the amplification of the 350 sequences, (II) and/or their complements and
 CC as oligomer probes for detecting the cytosine methylation state and/or
 CC single nucleotide polymorphisms (SNPs).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.

XX Sequence 5518 BP; 1383 A; 1452 G; 2541 T; 0 other;

Query Match 87.6%; Score 18.4; DB 24; Length 5518;
 Best Local Similarity 95.0%; Pred. No. 48;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTG 20
 |||||

Db 46 TAGGGGATTGGAGATTGTG 65

RESULT 4
 ID ABL32217 standard; DNA; 5518 BP.

AC ABL32217;
 XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 190.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antihemetic; antirheumatic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

XX 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043626.

XX (EPIC-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX Claim 1; SEQ ID NO 190; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 5518 BP; 1383 A; 1452 G; 2541 T; 0 other;

Query Match 87.6%; Score 18.4; DB 24; Length 5518;
 Best Local Similarity 95.0%; Pred. No. 48;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGTG 20
 |||||

Db 46 TAGGGGATTGGAGATTGTG 65

RESULT 5
 ID AAA51818 standard; DNA; 21 BP.

XX AAA51818;
 AC AAA51818;

XX 31-OCT-2000 (first entry)

DE Sense primer for CASP8 methylated-specific bisulfite treated DNA.

XX CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
 KW tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
 KW death receptor; apoptosis; cytosolic; gene therapy; primer; ss.

XX Homo sapiens.

OS WO200039347-A1.

XX 06-JUL-2000.

PF 30-DEC-1999; 99WO-US31280.

PR 31-DEC-1998; 98US-0114308.

PA (SUJD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Kidd VJ, Lahti JM, Teitz T;

XX WPI; 2000-452423/39.

XX Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
 PT prognosing cancer, comprises detecting a modification of genomic DNA
 PT comprising the CASP8 gene that results in inactivation of the gene
 XX Example 3; Page 61; 107pp; English.

XX Investigation of silencing of caspase-8 (CASP8) gene by methylation was
 CC analyzed using methylation-sensitive PCR analysis. Amplification in the
 CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
 CC mixtures containing bisulfite-treated DNA. Primers AAA51818-23 were
 CC designed to produce a 320 bp fragment in the upstream region of Casp8
 CC gene extending from nucleotides +221 to +541. Wild type primers were
 CC used to amplify the corresponding region of untreated genomic DNA.
 CC Controls without DNA were also performed. CASP8, a cysteine protease, is
 CC part of the death inducing signaling complex (DISC) associated with the
 CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
 CC tumour suppressor gene. The CASP8 promoter region sequences, in
 CC particular Region 1 and Region 2, are crucial to the design and execution
 CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
 CC Methylation PCR can be used to examine even minute amounts of patient
 CC material to demonstrate whether the CASP8 gene expresses an mRNA and
 CC protein product. The CASP8 gene has been localized to human chromosome
 CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
 CC treated by administering a vector that expresses a gene encoding
 CC functional CASP8 in cells. The cancer that is diagnosed or treated is a

PT comprising the CASP8 gene that results in inactivation of the gene
XX
PS Example 3, Page 61, 107pp; English.
CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was
CC analyzed using methylation-sensitive PCR analysis. Amplification of the
CC 5' untranslated region (UTR) of the CASP8 gene was performed in reaction
CC mixtures containing bisulfite treated DNA. Primers AAS1818-23 were
CC designed to produce a 320 bp fragment in the upstream region of CASP8
CC gene extending from nucleotides +221 to +541. Wild type primers were
CC used to amplify the corresponding region of untreated genomic DNA.
CC Controls without DNA were also performed. CASP8, a cysteine protease, is
CC part of the death inducing signaling complex (DISC) associated with the
CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
CC tumour suppressor gene. The CASP8 promoter region sequences, in
CC particular Region 1 and Region 2, are crucial to the design and execution
CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
CC Methylation PCR can be used to examine even minute amounts of patient
CC material to demonstrate whether the CASP8 gene expresses an mRNA and
CC protein product. The CASP8 gene has been localized to human chromosome
CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
CC treated by administering a vector that expresses a gene encoding
CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
CC tumour in which a myc gene is amplified, such as a neuroblastoma.
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
CC uterine cervical carcinoma can be diagnosed with the new method. A kit
CC for screening for a compound that induces death-receptor-mediated
CC apoptosis in cells containing an inactivated CASP8 gene is also
CC provided.
XX
SQ Sequence 21 BP; 5 A; 0 C; 9 G; 7 T; 0 other;
Query Match 100.0%; Score 21; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGGGATTGGAGATTGTGA 21
DB 1 TAGGGGATTGGAGATTGTGA 21
RESULT 2
ABL54318
ID ABL54318 standard; DNA; 5518 BP.
XX
AC ABL54318;
XX
DT 29-JUL-2002 (first entry)
XX
DE Chemically treated apoptosis gene complementary to gene #9.
XX
KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
KW amyotrophic lateral sclerosis; cancer; ds.
XX
OS Unidentified.
XX
FM WO200177164-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03969.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX

DR MPI; 2002-017444/02.
XX
PT Chemically modified sequences of genes associated with apoptosis are
PT useful to determine methylation patterns of genomic DNA samples for
PT diagnosis of associated diseases such as cancer
XX
PS Claim 1; Seq ID #18; 24pp; English.
XX
CC This invention relates to chemically pre-treated DNA of genes
CC associated with apoptosis. The nucleic acids are used to allocate
CC patients for specific therapy for HIV infection, Bloom syndrome,
CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
CC and cancers. This nucleotide sequence represents a chemically
CC treated apoptosis gene. Even SEQ ID numbers are the complementary
CC DNA strands to the odd SEQ ID numbers. The sequence data for this
CC patent is not represented in the printed specification but is based on
CC information supplied by the European patent office.
XX
SQ Sequence 5518 BP; 1383 A; 142 C; 1452 G; 2541 T; 0 other;
Query Match 87.6%; Score 18.4; DB 24; Length 5518;
Best Local Similarity 95.0%; Pred. No. 46;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TAGGGGATTGGAGATTGTG 20
DB 46 TAGGGGATTGGAGATTGTG 65
RESULT 3
ABN80019
ID ABN80019 standard; DNA; 5518 BP.
XX
AC ABN80019;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 36.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytosolic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07536.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR MPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a
PT sequence of a segment of chemically pretreated DNA of genes associated
PT with development
XX
PS Claim 1; SEQ ID NO 36; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 20.9513 Seconds
(without alignments)
2257.229 Million cell updates/sec

Title: US-09-477-082-31
Perfect score: 21
Sequence: 1 tagggatttgagattgtga 21

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_101002:*

1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	AAA51820	Sense primer for C
2	18.4	87.6	5518	AB154318	Chemically treated
3	18.4	87.6	5518	ABN80019	Human chemically m
4	18.4	87.6	5518	ABJ32217	Human immune syste
5	17.8	84.8	21	AAA51818	Sense primer for C
6	17.8	84.8	2204	AB122738	Drosophila melanog
7	17.8	84.8	2361	AB112786	Drosophila melanog
8	17.8	84.8	3536	AB113136	Drosophila melanog
9	17.8	84.8	13377	AAA6475	Tumour suppressor

10	17.8	84.8	13377	24	ABJ33462	Human immune syste
11	17	81.0	16914	24	AB170316	Chemically treated
12	17	81.0	16914	24	AA561254	Human gene regulat
13	16.8	80.0	898	21	AAC52475	Arabidopsis thalia
14	16.8	80.0	1199	23	AB111065	Drosophila melanog
15	16.8	80.0	2325	21	AAA57385	Pea piz2 gene 119n
16	16.8	80.0	3343	23	AB111064	Drosophila melanog
17	16.8	80.0	3441	21	AAA97382	Pea 119n-repressi
18	16.8	80.0	3916	22	AAK68032	Human immune/haema
19	16.8	80.0	6127	24	AB132592	Human immune syste
20	16.8	80.0	168575	22	AAH21613	Human hypoxerlin r
21	16.4	78.1	519	24	ABQ36984	Oligonucleotide fo
22	16.4	78.1	519	24	ABQ36985	Oligonucleotide fo
23	16.4	78.1	588	23	ABV55682	Human prostate exp
24	16.4	78.1	981	23	AA553751	Helicobacter pylor
25	16.4	78.1	11358	22	AAH22434	P450RAI-2 containi
26	16.2	77.1	422	22	AAC92173	Rabbit antiinflamag
27	16.2	77.1	452	22	AA541743	Genomic sequence #
28	16.2	77.1	452	22	AAK71554	Human immune/haema
29	16.2	77.1	452	22	AAK80621	Human immune/haema
30	16.2	77.1	640	24	ABQ37686	Oligonucleotide fo
31	16.2	77.1	640	24	ABQ37687	Oligonucleotide fo
32	16.2	77.1	810	23	ABV10247	Human prostate exp
33	16.2	77.1	1587	21	AAK42056	Arabidopsis thalia
34	16.2	77.1	3227	22	AAK44254	Human immune/haema
35	16.2	77.1	5310	24	ABK33934	Human DNA for stag
36	16.2	77.1	6408	24	ABL33120	Human immune syste
37	16.2	77.1	7763	24	AB134437	Human immune syste
38	16.2	77.1	9927	24	AB132113	Human immune syste
39	16.2	77.1	11046	24	ABK31536	Signal transductio
40	16.2	77.1	11996	24	ABL34493	Human metastasis a
41	16.2	77.1	15765	22	ABAI7911	Human nervous syst
42	16.2	77.1	15765	22	ABAI8250	Human nervous syst
43	16.2	77.1	15765	22	ABAI8276	Human nervous syst
44	16.2	77.1	15765	22	ABAI9009	Human nervous syst
45	16.2	77.1	16772	22	ABA20862	Human nervous syst

ALIGNMENTS

RESULT 1						
ID	AAA51820	standard:	DNA:	21	BP.	
XX	AAA51820:					
AC	31-OCT-2000	(first entry)				
XX						
DT						
DE	Sense primer for CASP8 unmethylated-specific disulfite treated DNA.					
XX						
XX	CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;					
KW	tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;					
KW	death receptor; apoptosis; cytostatic; gene therapy; primer; ss.					
OS	Homo sapiens.					
XX						
PN	WO200039347-A1.					
XX						
PD	06-JUL-2000.					
XX						
PF	30-DEC-1999:	99WO-US31280.				
XX						
PR	31-DEC-1998:	98US-0114308.				
XX						
PA	(SUD-) ST JUDE CHILDREN'S RES HOSPITAL.					
XX						
PI	Kidd VJ, Lahti JM, Teltz T;					
XX						
DR	WPI; 2000-452423/39.					
XX						
PT	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or					
PT	prognosing cancer, comprises detecting a modification of genomic DNA					

```

repeat_region 19590..20045
/note="LIM4 repeat: matches 2859..3326 of consensus"
repeat_region 20099..20355
/note="LIM4 repeat: matches 3603..3874 of consensus"
misc_feature 20275..20787
/note="match: GSS: Em:AQ547282"
repeat_region 21667..22029
/note="LIM1B repeat: matches 4..375 of consensus"
repeat_region 22560..22646
/note="L2 repeat: matches 2624..2710 of consensus"
repeat_region 23310..23954
/note="L2 repeat: matches 1993..2643 of consensus"
misc_feature complement(23742..24153)
/note="match: GSS: Em:AQ182243"
misc_feature 24153..24603
/note="match: GSS: Em:AQ880736"
misc_feature 24173..24662
/note="match: GSS: Em:AQ498082"
repeat_region 26631..26684
/note="27 copies 2 mer aa 74% conserved"
repeat_region 26986..27049
/note="32 copies 2 mer tt 67% conserved"
repeat_region 27058..28403
/note="LIP1A16 repeat: matches 4771..6156 of consensus"
repeat_region 28404..28699
/note="LIM4 repeat: matches 6..304 of consensus"
repeat_region 28700..29162
/note="LIP1A16 repeat: matches 4315..4771 of consensus"
repeat_region 29948..30441
/note="LIM1D repeat: matches 4..505 of consensus"
misc_feature complement(30606..31171)
/note="match: GSS: Em:AQ280648"
repeat_region 30764..30992
/note="L1 repeat: matches 4435..4668 of consensus"
repeat_region 31560..33433
/note="LIM9 repeat: matches 4436..6306 of consensus"
repeat_region 33796..34254
/note="L2 repeat: matches 2073..2570 of consensus"
repeat_region 34418..34526
/note="MIR repeat: matches 31..139 of consensus"
repeat_region 34653..34951
/note="LIM1 repeat: matches 5827..6105 of consensus"
repeat_region 34952..35443
/note="LOR1 repeat: matches 1..497 of consensus"
repeat_region 35444..35913
/note="LIM1 repeat: matches 5285..5827 of consensus"
repeat_region 35955..36631
/note="LIM1 repeat: matches 4545..5298 of consensus"
repeat_region 36633..37090
/note="LIM4 repeat: matches 3548..4026 of consensus"
repeat_region 37269..37376
/note="LIM4 repeat: matches 3131..3242 of consensus"
repeat_region 37510..37705
/note="LIMC repeat: matches 1813..2005 of consensus"
repeat_region 37760..38650
/note="LIM6 repeat: matches 5231..6088 of consensus"
repeat_region 38649..38823
/note="L1 repeat: matches 4553..4718 of consensus"
repeat_region 38824..39051
/note="PTR3 repeat: matches 743..965 of consensus"
repeat_region 39131..39181
/note="LTR30 repeat: matches 1..52 of consensus"
repeat_region 39185..39292
/note="PTR5 repeat: matches 1570..1679 of consensus"
repeat_region 39293..39981
/note="LTR12 repeat: matches 1..671 of consensus"
repeat_region 39982..40184
/note="L1 repeat: matches 4715..4912 of consensus"
repeat_region 40338..40650
/note="ALUX repeat: matches 1..306 of consensus"
repeat_region 40866..41242
/note="LTR16A repeat: matches 61..445 of consensus"
repeat_region 41259..41576

```

```

repeat_region /note="L2 repeat: matches 1777..2128 of consensus"
41824..41926
/note="LIP1A16 repeat: matches 6041..6143 of consensus"
repeat_region 42069..42189
/note="PIAM.C repeat: matches 7..126 of consensus"
misc_feature 43350..43736
/note="match: GSS: Em:B60315"
misc_feature 43373..43929
/note="match: GSS: Em:AQ269706"
misc_feature 44214..44643
/note="match: GSS: Em:AQ358471"
misc_feature 44232..44969

```

Query Match 84.8%; Score 17.8; DB 9; Length 98368;
 Best Local Similarity 90.5%; Pred. No. 2.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGGGATTGTGA 21
 Db 50285 TAGGGATTGGATTGTGTA 50305

Search completed: July 8, 2003, 04:02:46
 Job time : 79.7495 secs

19253..19416,19674..19873,19922..19960))
 /gene="Y39DB8.3"
 /ecodon_start=1
 /product="Hypothetical protein Y39DB8.3"
 /protein_id="AA06053.1"
 /db_xref="GI:7331845"
 /translation="MTAVETIKKQNDITLIRBEDVADVDDDDMERKYNITLTH
 GNTTITISDEARDLTKHMDQSLAGLGAATVAVTINKKSKSEKEHTYCIASK
 NYTAHAKCVTVLDQYKRNKLEKSTASLKLHOKQSDGALKSNMNEYEY
 KSGGEFVYKSGSGFVDEMRKROAVIREVOTRKSEYELDNKSLPLIARKLT
 ELYRAGKNEKPEKQKMODVIOEIKNSTRICKSONKEMKRFSEKFTIMKGLNP
 DKAMOSIGMEDLFADEPILSEKEDAKDKAKEMSPDKILNEPILIRAIKIGMAA
 GTRKDAEMSDKKIALISQFMSITLDEPVANNTISLSITSLIGESGDDROSLT
 KALKLESTGQERMNWVLEASGVETVANKRAKEEKEERKMDQFVDRKGPFTS
 KENATQIGVEYAGKLELDLDFPKSLSSQRTPKWKQLGYHRNSRPFQNFENSKIF
 GISRKYQRCHEKLFPSQKFEPTPKFFPNFKIFQMESKNKGYTVMNDKQLETTIYGH
 GSPFNSEALQKFRGISPSMPARIEKRMIAPEEMKFDISRYKIE"

BASE COUNT 6587 a 3794 c 3934 g 5988 t
 Query Match 84.8%; Score 17.8; DB 3; Length 20303;
 Best Local Similarity 90.5%; Pred. No. 2e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGGAGATTGCA 21
 1 ||||||||||||| |||| ||
 DB 6210 TAGGGATTGGAGATTGAGA 6230

RESULT 15
 AL136453 98368 bp DNA linear PRI 31-MAR-2000
 LOCUS AL136453
 DEFINITION Human DNA sequence from clone RP1-249K20 on chromosome 10. Contains
 STS and GSSs, complete sequence.
 ACCESSION AL136453
 VERSION AL136453.3 GI:6912131
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 98368)
 Graffham D.
 Direct Submission
 Submitted (14-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Feb 7, 2000 this sequence version replaced gi:67233337.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence has been finished according to sequence map criteria
 as follows. An attempt is made to resolve all sequencing problems,
 such as compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/Celegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Ch10
 RP1-249K20 is from the library RPCT-1 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see http://bacpac.med.buffalo.edu/
 VECTOR: pCYPAC2

FEATURES
 source
 This sequence is the entire insert of clone RP1-249K20 The true
 left end of clone RP4-53118 is at 61968 in this sequence.

1..98368
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP1-249K20"
 /clone_lib="RPCT-1"
 /complement(136..800)
 /note="match: GSS: Em:AQ779626"
 154..858
 /note="match: GSS: Em:AQ009393"
 164..868
 /note="match: GSS: Em:AQ009392"
 208..592
 /note="match: GSS: Em:B86219"
 /complement(364..800)
 /note="match: GSS: Em:AQ465238"
 2971..2996
 /note="13 copies 2 mer aa 100% conserved"
 3274..3309
 /note="9 copies 4 mer caaa 97% conserved"
 4106..4317
 /note="MIR repeat: matches 15..226 of consensus"
 7363..7540
 /note="LIM1 repeat: matches 5381..5563 of consensus"
 /complement(7577..7983)
 /note="match: GSS: Em:AQ755541"
 7606..7798
 /note="MIR repeat: matches 2..192 of consensus"
 7880..7956
 /note="L2 repeat: matches 2666..2747 of consensus"
 9022..9252
 /note="AluX repeat: matches 1..231 of consensus"
 9089..9457
 /note="match: STS: Em:LL17677"
 9279..9400
 /note="61 copies 2 mer aa 63% conserved"
 9284..9371
 /note="8 copies 11 mer 68% conserved"
 10062..10217
 /note="MIR repeat: matches 57..221 of consensus"
 11677..11813
 /note="Alu repeat: matches 165..301 of consensus"
 12586..12866
 /note="LIM4 repeat: matches 4856..5160 of consensus"
 12899..13407
 /note="LIM4 repeat: matches 4886..5403 of consensus"
 13271..13789
 /note="match: GSS: Em:AQ270786"
 13417..13645
 /note="LIM4 repeat: matches 6043..6289 of consensus"
 13736..14644
 /note="LIM4 repeat: matches 5417..6288 of consensus"
 14365..14712
 /note="match: STS: Em:Z53533"
 14732..15046
 /note="MER34 repeat: matches 190..534 of consensus"
 15047..16331
 /note="LIM4 repeat: matches 514..1735 of consensus"
 16356..16583
 /note="LIM4 repeat: matches 3909..4135 of consensus"
 17777..17838
 /note="MER63 repeat: matches 979..1046 of consensus"
 18114..18346
 /note="MIR repeat: matches 11..249 of consensus"
 /complement(18285..18763)
 /note="match: STS: Em:G63247"
 18450..19006
 /note="match: GSS: Em:AQ87926"
 19193..19386
 /note="LIMC repeat: matches 952..1353 of consensus"

VERSION ACO06749.1 GI:4263189
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 20303)
 AUTHORS Waterston, R.
 TITLE Genome sequence of the nematode *C. elegans*: a platform for
 investigating biology. The *C. elegans* Sequencing Consortium
 JOURNAL Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613
 PUBMED 9851916
 REFERENCE 2 (bases 1 to 20303)
 AUTHORS Becker, M., Graves, T. and Yeakum, M.
 TITLE The sequence of *C. elegans* cosmid Y39DB8
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 20303)
 AUTHORS Waterston, R. H.
 TITLE Direct Submission
 JOURNAL Submitted (23-FEB-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 4 (bases 1 to 20303)
 AUTHORS Waterston, R. H.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 REFERENCE 5 (bases 1 to 20303)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (28-MAR-2000) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 6 (bases 1 to 20303)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (24-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 COMMENT Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: r.w@nematoe.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is T22H9, 3951 bp overlap; the 3' cosmid is 2K6, 200
 bp overlap. Actual start of this cosmid is at base position 31064
 of Y56G88; actual end is at 27075 of F48G7.

NOTES:

Coding sequences below are the result of integration and manual

review of the following data : computer analysis using the program
 GeneFinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yui Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The *C.*
elegans ORFome cloning project (<http://wofdb.dcfi.harvard.edu/>),
 similarity to other proteins from Blastx analyses
 (<http://blast.wustl.edu/>), sequence conservation with *C. briggsae*
 using Jim Kent's MABA alignment program (Genome Research
 10:1115-1125, 2000), individual *C. elegans* Genbank submissions,
 and personal communications with *C. elegans* researchers. TRANS
 are predicted using the program TRANScan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).

Location/Qualifiers

1. 20303
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="v"
 /clone="Y39DB8"
 5238..8372
 /gene="Y39DB8.2"
 /note="for a graphical representation of this gene see:
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8.2;class=Sequence>"
 join(5238..5295,5343..5596,5649..5845,6631..6794,
 6844..7138,7183..7333,7462..7716,7763..8101,8157..8372)
 /gene="Y39DB8.2"
 /codon_start=1
 /product="Hypothetical protein Y39DB8.2"
 /protein_id="AAAF60532.1"
 /db_xref="GI:7331844"
 /translation="MGFTFRLFEVYFVGVRSVSGSIISREVPVTSPPQOSTLL
 PFWLWENKPEFTSKPKRNALREGLIDQVRGKIPKELAPVNTKIDENKKE
 ODTPEITRDEPDVADPPDDVHRVNTLTDGKNTITITIDEPADDLVKKWMSGLA
 GLMGAVTVTVIRKMKTKSEKDRHICINESKNTYAAKCYVYVLDQNNKNEKLKY
 STLATLRSHRKTRADNDITKTRLDHDKTVKSGASEYRVRKRAGESIFDEMKRQ
 AAIRSNVQTRKSYTIRKNSLSPALILARKILELVRAGNKOEPKRWQVIOEIKKE
 SPRIGKNNKNNKMRKRSKYSTVSTKSGKGLNRLNTMGEDLFEDEPIISEMIL
 KKRKMEERKNSDDDKIMEPIKILIREAKIGVAMAGKSGEIGDKVALLSPORMSL
 PDEVANDTSLSPSILSHGSGTDESLSTKRLPLMEDGOEKNFYLEASGVT
 ETVEKMRITAEKEKDEDEPRKDFVCDGSGAPFTKENTVEITGEVZATIDILINGYKT
 MSKDOLDALNSGYSIMDDKQDMLXGSGSPYNNSEALDKRGITLSEMPRIODIR
 LLAREDKTELSKRVRLI"
 10676..11343
 /gene="Y39DB8.1"
 /note="for a graphical representation of this gene see:
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8.1;class=Sequence>"
 join(10676..10709,11142..11472,11529..12042,12415..12517,
 12562..12680,12730..12900,12963..13130,13280..13343)
 /gene="Y39DB8.1"
 /codon_start=1
 /product="Hypothetical protein Y39DB8.1"
 /protein_id="AAAF60531.1"
 /db_xref="GI:7331843"
 /translation="MNFTEGAVYGLDNTSSSDTQSVRLKHQLTETGTPSKLPLFLN
 LIAVERKSHNGSLQFTIVDDYKRGIGSDPVLAINMTDGNKTVKISEEAYQL
 IAKVWDQASGFMALMSYTKLDKRLDEPRHILQHNKNOVAKSITFHAQCVVTLIKRL
 DQQRIRIKRKLKFRSOLRRARAKRLPEYRDRKRVQKOOVLLKKKEKSTP
 FSLAKTLTDIYRNNKREKPKKNQVYINIDELNRIKOKKKQDQDFNRFKYS
 RPKMKIGINPRTAFRAGIDSGEISANTGKTSEREOVYKRLMIRGQYKIGVM
 LTGQVSNFDDKRIALMSQFMSVLPDEQANDTVLLSPVFALHESSELDOMTSL
 RAKRFLENGHDDMMNPFVEASGVTEAEVKEFDRKQEMEFKRFHSENGQVET
 RQNSSEMVGDTYSKIDSMEKLRKMSAAMHNNKSTGYAIVMSQOISEFPGSPYN
 DSHAVENYENLRNDIPDLENNIHQAREEQAFVAROVGHELMRTTHYKNCISRK
 M"
 complement(15523..19960)
 /gene="Y39DB8.3"
 /note="for a graphical representation of this gene see:
<http://www.wormbase.org/db/seq/sequence?name=Y39DB8.3;class=Sequence>"
 complement(join(15523..15730,15925..16058,16137..16475,
 17187..17311,17828..17954,18005..18155,18905..19199,

```

misc_feature      /note="assembly_fragment"
                  120974..136795
misc_feature      /note="assembly_fragment"
                  136896..151116
misc_feature      /note="assembly_fragment"
                  151217..179686
BASE COUNT      58041 a 31435 c 31392 g 57013 t 1805 others
ORIGIN
Query Match      85.7%; Score 18; DB 2; Length 179686;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY
1 TAGGGATTGGAGATTG 18
|||||
6412 TAGGGATTGGAGATTG 6395

RESULT 11
LOCUS            AK096637      4369 bp      mRNA      linear      PRI 15-JUL-2002
DEFINITION      Homo sapiens cDNA FLJ39318 f1s, clone OCBFP2014089.
ACCESSION       AK096637
VERSION          AK096637.1 GI:21756173
KEYWORDS         cllgo capping; f1s (full insert sequence);
SOURCE           Homo sapiens fetal brain cDNA to mRNA, clone_l1b:OCBPF2
                  clone:OCBFP2014089.
ORGANISM         Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS          1
                  Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
                  Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
                  Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saico,K., Nishikawa,T.,
                  Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
                  Kikuchi,H., Kanda,K., Wagatsuma,M., Kurakawa,K., Kanehori,K.,
                  Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
                  Sugano,S., Nagahara,K., Masuhio,Y., Nagai,K. and Isogai,T.
                  NEDO human cDNA sequencing project
TITLE            Unpublished
JOURNAL          2 (bases 1 to 4369)
REFERENCE
AUTHORS          Isogai,T. and Yamamoto,J.
TITLE            Direct Submission
JOURNAL          Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
                  Kazusa-Kamatari, Kisarazu, Chiba 292-0912, Japan
                  (E-mail:genomise@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT          NEDO human cDNA sequencing project supported by Ministry of
                  Economy, Trade and Industry of Japan; cDNA full insert sequencing:
                  Research Association for Biotechnology (RAB); cDNA library
                  construction: Helix Research Institute (HRI) (supported by Japan
                  Key Technology Center etc.); 5'- & 3'-end one pass sequencing; RAB,
                  HRI, and Biotechnology Center, National Institute of Technology and
                  Evaluation; clone selection for full insert sequencing: HRI and
                  RAB; annotation: HRI and RAB.
FEATURES
Source
1. 4369
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBFP2014089"
/tissue_type="brain"
/clone_l1b="OCBPF2"
/dev_stage="fetal"
/note="cloning vector: PME18SFL3"
BASE COUNT      1163 a 920 c 1044 g 1242 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 9; Length 4369;
Best Local Similarity 90.3%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY
1 TAGGGATTGGAGATTGTA 21
|||||

```

```

Db 1369 TAGGCCTTTGGAGATTGTA 1389

RESULT 12
AX251229
LOCUS            AX251229      13377 bp      DNA      linear      PAT 05-OCT-2001
DEFINITION      Sequence 197 from Patent WO018912.
ACCESSION       AX251229
VERSION          AX251229.1 GI:15984652
KEYWORDS         synthetic construct.
SOURCE           synthetic construct.
ORGANISM         artificial sequences.
REFERENCE
AUTHORS          1 (bases 1 to 13377)
                  Olek,A., Piepenbrock,C. and Berlin,K.
TITLE            Diagnosis of diseases associated with tumor suppressor genes and
                  oncogenes
JOURNAL          Patent: WO 0168912-A 197 20-SEP-2001;
                  Epigenomics AG (DE)
FEATURES
Source
1. 13377
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3915 a 146 c 3046 g 6270 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 6; Length 13377;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY
1 TAGGGATTGGAGATTGTA 21
|||||
1704 TAGGGATTGGAGATTGTA 1724

RESULT 13
AX346364
LOCUS            AX346364      13377 bp      DNA      linear      PAT 01-FEB-2002
DEFINITION      Sequence 1435 from Patent WO0200928.
ACCESSION       AX346364
VERSION          AX346364.1 GI:18494250
KEYWORDS         synthetic construct.
SOURCE           synthetic construct.
ORGANISM         artificial sequences.
REFERENCE
AUTHORS          1
                  Olek,A., Piepenbrock,C. and Berlin,K.
TITLE            Diagnosis of diseases associated with the immune system
JOURNAL          Patent: WO 0200928-A 1435 03-JAN-2002;
                  Epigenomics AG (DE)
FEATURES
Source
1. 13377
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT      3915 a 146 c 3046 g 6270 t
ORIGIN
Query Match      84.8%; Score 17.8; DB 6; Length 13377;
Best Local Similarity 90.5%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY
1 TAGGGATTGGAGATTGTA 21
|||||
1704 TAGGGATTGGAGATTGTA 1724

RESULT 14
AC006749
LOCUS            AC006749      20303 bp      DNA      linear      INV 24-MAY-2002
DEFINITION      Caenorhabditis elegans cosmid y39DB8, complete sequence.
ACCESSION       AC006749

```

TITLE
JOURNAL
REFERENCE
AUTHORS

Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
Sever, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Tirrell, A.,
Travers, M., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
Zody, M.

Direct Submission
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
(bases 1 to 179686)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abrahm, H., Allen, N.,
Anderson, S., Baldwin, J., Bara, N., Bastien, V., Beda, F.,
Boguslavsky, L., Bouckhagalter, B., Brown, A., Burkett, G.,
Campoliano, A., Castle, A., Choeppel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,
Dodg, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, M., Gage, D.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Holland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lenoczky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McKean, P., McGuck, A., McKernan, K., McPheters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Navlor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Sever, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Testfaye, S., Theodore, J., Tirrell, A., Travers, M., Trifilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 4, 2000 this sequence version replaced g1:7229919.
All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: 16904

Center clone name: 390G_24

----- Summary Statistics

Sequencing vector: M13, M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 16279 bases at least Q40

Consensus quality: 170959 bases at least Q30

Consensus quality: 174650 bases at least Q20

Insert size: 176000; agarose-fp

Insert size: 177866; sum-of-ctnigs

Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 3.9 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1169: contig of 1169 bp in length
* 1170 1269: gap of 100 bp
* 1270 2277: contig of 1008 bp in length
* 2278 2377: gap of 100 bp
* 2378 3411: contig of 1034 bp in length
* 3412 3511: gap of 100 bp
* 3512 5019: contig of 1508 bp in length

FEATURES

source

* 5020 5119: gap of 100 bp
* 5120 7487: contig of 2368 bp in length
* 7488 7587: gap of 100 bp
* 7588 9310: contig of 1733 bp in length
* 9311 9410: gap of 100 bp
* 9411 15238: contig of 5828 bp in length
* 15239 15338: gap of 100 bp
* 15339 25751: contig of 10413 bp in length
* 25752 25852: gap of 100 bp
* 25853 33183: contig of 7332 bp in length
* 33184 33283: gap of 100 bp
* 33284 44080: contig of 10797 bp in length
* 44081 44180: gap of 100 bp
* 44181 54508: contig of 10528 bp in length
* 54509 54608: gap of 100 bp
* 54609 67237: contig of 12629 bp in length
* 67238 67337: gap of 100 bp
* 67338 79729: contig of 12392 bp in length
* 79730 79829: gap of 100 bp
* 79830 96089: contig of 16260 bp in length
* 96090 96189: gap of 100 bp
* 96190 108632: contig of 12443 bp in length
* 108633 108732: gap of 100 bp
* 108733 120873: contig of 12141 bp in length
* 120874 120973: gap of 100 bp
* 120974 136795: contig of 15822 bp in length
* 136796 136895: gap of 100 bp
* 136896 151116: contig of 14221 bp in length
* 151117 151216: gap of 100 bp
* 151217 179686: contig of 28470 bp in length.
Location/Qualifiers
1. 179686
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/map="5"
/clone="RP11-390G24"
/clone_1lb="RPCT-11 Human Male BAC"
1. 1169
/note="assembly-fragment"
1270. 2277
/note="assembly-fragment"
2378. 3411
/note="assembly-fragment"
3512. 5019
/note="assembly-fragment"
5120. 7487
/note="assembly-fragment"
7588. 9310
/note="assembly-fragment"
9411. 15238
/note="assembly-fragment"
15339. 25751
/note="assembly-fragment"
25852. 33183
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
33284. 44080
/note="assembly-fragment"
44181. 54508
/note="assembly-fragment"
clone_end:IT7
vector_side:right"
54609. 67237
/note="assembly-fragment"
67338. 79729
/note="assembly-fragment"
79830. 96089
/note="assembly-fragment"
96190. 108632
/note="assembly-fragment"
108733. 120873

```

*      81943      82042: gap of unknown length
*      82043      150022: contig of 67980 bp in length
*      150023      150122: gap of unknown length
*      150123      152401: contig of 2279 bp in length
*      152402      152501: gap of unknown length
*      152502      158391: contig of 5890 bp in length
*      158392      158491: gap of unknown length
*      158492      162895: contig of 4404 bp in length
*      162896      162995: gap of unknown length
*      162996      168149: contig of 5154 bp in length
*      168150      168249: gap of unknown length
*      168250      176144: contig of 7895 bp in length
*      176145      176244: gap of unknown length
*      176245      186417: contig of 10173 bp in length
*      186418      186517: gap of unknown length
*      186518      199541: contig of 13024 bp in length
*      199542      199642: gap of unknown length
*      199642      214787: contig of 15146 bp in length.
FEATURES
source      Location/Qualifiers
            1..214787
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="X"
            /clone="RP23-326J1"
            1..15958
            /note="assembly_name:Contig10"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            1..15958
            /note="assembly_name:Contig11"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            47260..81942
            /note="assembly_name:Contig12"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            82043..150022
            /note="assembly_name:Contig13"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            150123..152401
            /note="assembly_name:Contig2"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            152502..158391
            /note="assembly_name:Contig3"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            158492..162895
            /note="assembly_name:Contig4"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            162996..168149
            /note="assembly_name:Contig5"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            168250..176144
            /note="assembly_name:Contig6"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            176245..186417
            /note="assembly_name:Contig7"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            186518..199541
            /note="assembly_name:Contig8"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            199642..214787
            /note="assembly_name:Contig9"
            /db_xref="taxon:10090"
            /clone="RP23-326J1"
            57078 a 49921 c 49241 g 57431 t 1116 others
BASE COUNT
ORIGIN
Query Match      87.6%; Score 18.4; DB 2; Length 214787;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      2 AGGGGATTGGAGATTGCA 21
Db      109370 AGGGGACTTGGAGATTGCA 109389
RESULT 9
LOCUS      AC091938      70122 bp      DNA      linear      PRI 31-MAY-2002
DEFINITION      Homo sapiens chromosome 5 clone RP11-329N11, complete sequence.
ACCESSION      AC091938
VERSION      AC091938.2      GI:21281476
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 70122)
DOE Joint Genome Institute and Stanford Human Genome Center.

```

```

TITLE      Direct Submission
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 70122)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
              3 (bases 1 to 70122)
REFERENCE    DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS      Submitted (31-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
              On May 31, 2002 this sequence version replaced gi:14333874.
COMMENT      Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 99.5% of Sequence.
              Estimated Total Number of Errors is 0.2.
              NOTE: This insert is not the entire sequence of the clone (entire
              sequence is 153kb). It is clipped at the overlaps with AC022139 and
              AC109455. The number of bases overlapped with AC022139 is 12649 and
              with AC109455 is 3639.
FEATURES
source      Location/Qualifiers
            1..70122
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-329N11"
BASE COUNT      23470 a 11786 c 12143 g 22723 t
ORIGIN
Query Match      85.7%; Score 18; DB 9; Length 70122;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 TAGGGGATTGGAGATTG 18
Db      27627 TAGGGGATTGGAGATTG 27644
RESULT 10
LOCUS      AC023987/c      179686 bp      DNA      linear      HTG 24-AUG-2002
DEFINITION      Homo sapiens chromosome 5 clone RP11-390G24 map 5, WORKING DRAFT
ACCESSION      AC023987
VERSION      AC023987.3      GI:7705153
KEYWORDS      HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 179686)
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE      Homo sapiens chromosome 5, clone RP11-390G24
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 179686)
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
              Anderson, S., Baldwin, J., Barna, N., Bedalov, F., Boguslavsky, L.,
              Boukhalter, B., Brown, A., Burkett, G., Campoliano, A., Castle, A.,
              Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
              DeRellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
              Fenesstor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
              Galagan, J., Gargyala, S., Glade, S., Goyette, M., Graham, L.,
              Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
              Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
              Klein, J., Landers, T., Laroque, K., Lehotzky, J., Levine, R.,
              Lieu, C., Liu, G., Locke, K., MacDonald, P., Marguis, N., McCarthy, M.,
              McKean, P., McGurk, A., McKernan, K., McPheters, R., Melidim, J.,
              Menkus, L., Mihova, T., Miranda, C., Menga, V., Morrow, J., Naylor, J.,
              Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M.,
              Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,

```

```

repeat_region complement(4063, .4146)
/rpt_family="MIR"
repeat_region complement(4701, .4881)
/rpt_family="MIR"
repeat_region 4959, .4986
/rpt_family="AT-rich"
repeat_region 5136, .5373
/rpt_family="MLTII"
repeat_region complement(5374, .5639)
/rpt_family="AluSx"
repeat_region 5640, .5876
/rpt_family="MTII"
repeat_region complement(6582, .6683)
/rpt_family="MIR"
repeat_region 7263, .7329
/rpt_family="Charlie8"
repeat_region complement(7333, .7638)
/rpt_family="AluSq"
repeat_region 7723, .7847
/rpt_family="Charlie8"
repeat_region 7940, .8083
/rpt_family="Charlie8"
repeat_region complement(8330, .8639)
/rpt_family="AluSx"
repeat_region complement(8649, .8683)
note="Single clone coverage"
8844, .9152
/rpt_family="AluY"
repeat_region 9242, .9401
/rpt_family="L1PBI"
repeat_region 9402, .9586
/rpt_family="AluY"
repeat_region 9587, .9664
/rpt_family="AluSp/q"
repeat_region complement(9914, .10340)
/rpt_family="L2"
repeat_region 10749, .10878
/rpt_family="Charlie8"
repeat_region complement(11241, .11322)
/rpt_family="L2"
repeat_region 11651, .11959
/rpt_family="AluSx"
repeat_region 12129, .12154
/rpt_family="AT-rich"
repeat_region complement(12321, .12638)
/rpt_family="AluSg"
repeat_region complement(13500, .13642)
/rpt_family="MIR3"
repeat_region complement(13659, .13955)
/rpt_family="AluSp"
repeat_region complement(13973, .14204)
/rpt_family="MIR"
repeat_region 14345, .14369
/rpt_family="(CAAA)n"
repeat_region 14693, .14753
/rpt_family="(TTTA)n"
repeat_region complement(14758, .15037)
/rpt_family="AluSx"
repeat_region 17097, .17142
/rpt_family="(CA)n"
repeat_region 17949, .18097
/rpt_family="L1ME4A"
repeat_region complement(18645, .18755)
/rpt_family="MIR"
repeat_region 18982, .19069
/rpt_family="MIR"
repeat_region 19071, .19211
/rpt_family="FLAM_C"
repeat_region 19214, .19339
/rpt_family="MIR"
repeat_region complement(19411, .19606)
/rpt_family="MIR"
repeat_region 19725, .19777

```

```

repeat_region /rpt_family="MIR"
20135, .20247
/rpt_family="MIR3"
repeat_region complement(20248, .20536)
/rpt_family="AluSx"
repeat_region 20537, .20556

Query Match 87.6% Score 18.4; DB 9; Length 188868;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGGGATTGGAGATTGTGA 21
|||||
Db 2388 AGGGGATTGGAGATTGTGA 2369

RESULT 8
AC123853 214787 bp DNA linear HTG 01-JUN-2002
LOCUS Mus musculus chromosome UNK clone RP23-326J1, WORKING DRAFT
DEFINITION
ACCESSION AC123853
VERSION AC123853.1 GI:21307546
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 214787)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214787)
AUTHORS McPherson, J.D. and Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
Project Information
Center project name: M_BA0326J01
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 207948 bases at least Q40
Consensus quality: 208444 bases at least Q30
Consensus quality: 208788 bases at least Q20
Insert size: 171000; agarose-ftp
Insert size: 213687; sum-of-ctngs
Quality coverage: 11.36 in Q20 bases; agarose-ftp
Quality coverage: 9.45 in Q20 bases; sum-of-ctngs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 15958: contig of 15958 bp in length
* 15959 16038: gap of unknown length
* 16059 47159: contig of 31101 bp in length
* 47160 47259: gap of unknown length
* 47260 81942: contig of 34683 bp in length

```

```

misc.feature      /note="assembly_name:Contig3"
                  136514..140443
misc.feature      /note="assembly_name:Contig4"
                  140544..148908
misc.feature      /note="assembly_name:Contig6"
                  149009..160432
misc.feature      /note="assembly_name:Contig7"
                  160533..171996
misc.feature      /note="assembly_name:Contig8"
                  172097..186048
misc.feature      /note="assembly_name:Contig9"
                  50542 a 44621 c 41477 g 48405 t 1003 others
BASE COUNT
ORIGIN
Query Match      87.6% Score 18.4; DB 2; Length 186048;
Best Local Similarity 95.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 2 AGGGGATTTGGAGATTTGTGA 21
Db 72797 AGGGGATTTGAAGATTTGTGA 72778

```

```

RESULT 7
AC023232/c      188868 bp      DNA      linear      PRI 01-JUL-2002
LOCUS      Homo sapiens chromosome 11, clone RP11-793111, complete sequence.
AC023232
VERSION      AC023232.16 GI:21637520
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
  1 (bases 1 to 188868)
  Birren,B., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome 11, clone RP11-793111
  Unpublished
  2 (bases 1 to 188868)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
  Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
  Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
  Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
  Dearellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
  Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
  Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
  Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
  Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
  Macdonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K.,
  McSheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
  Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
  Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
  Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
  Stojanovic,N., Subramanian,A., Talamas,J., Testfaye,S., Theodore,J.,
  Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.T.,
  Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  3 (bases 1 to 188868)
  Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
  Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
  Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
  Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
  Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
  Fato,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
  Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
  Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
  Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,J., LaRocque,K.,
  Llanazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
  Liu,G., Maclean,C., Macdonald,P., Major,J., Marguis,N.,
  Matthews,C., McCarthy,W., McEwan,P., McKernan,K., Meldrim,J.,
  Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J., Nguyen,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
  4 (bases 1 to 188868)
  Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
  Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
  Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
  Cooke,P., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
  Fato,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D., Galagan,J.,
  Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
  Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
  Karatas,A., Kells,J., Landers,T., Levine,R., Lintblad-Toh,K.,
  Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
  McCarthy,W., Meldrim,J., Meneus,L., Mihova,T., Mlenka,V.,
  Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
  O'Connor,T., O'Donnell,P., Oliver,T., Oliver,J., Peterson,K.,
  Phunkhang,P., Pierre,N., Raymond,C., Reta,R., Rise,C., Rogov,P.,
  Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
  Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
  Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
  Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
  Zembek,L., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Jul 1, 2002 this sequence version replaced gi:21591935.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  -----Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence.submissions@genome.wi.mit.edu
  -----Project Information
  Center project name: L6735
  Center clone name: 793_111
  -----
  Location/Qualifiers
  1..188868
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="11"
  /map="11"
  /clone="RP11-793111"
  /clone_idb="RPC1-11 Human Male BAC"
  585..639
  /rpt_family="(A)n"
  1098..1295
  /rpt_family="MIR"
  1365..1584
  /rpt_family="MIR"
  2471..2546
  /rpt_family="L2"
  2700..2848
  /rpt_family="L2"
  2860..3079
  /rpt_family="L2"
  3214..3347
  /rpt_family="FLAM_C"
  complement(3363..3677)
  /rpt_family="AlusP"

```


JOURNAL Submitted (20-APR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 20, 2002 this sequence version replaced g1:17149713.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L21818

Center clone name: 2183_M_20

Location/Qualifiers

1. 136372

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="11"

/clone="CIRD1-2183M20"

/clone_lib="CIRD1 Human BAC"

complement(597..704)

/rpt_family="MIR"

complement(1858..2091)

/rpt_family="MIR"

complement(2120..2175)

/rpt_family="MIR"

complement(2353..2654)

/rpt_family="AluSx"

4234..4365

/rpt_family="MIR3"

complement(4887..4953)

/rpt_family="MIR"

5349..5395

/rpt_family="MIR"

5420..5705

/rpt_family="AluSx"

5706..5726

/rpt_family="AT-rich"

complement(6150..6222)

/rpt_family="Charlie1"

6267..6564

/rpt_family="AluSg"

complement(6569..6974)

/rpt_family="Charlie1"

complement(7026..7042)

/rpt_family="Charlie1"

7043..7422

/rpt_family="MER101"

7423..7732

/rpt_family="AluSx"

7733..7866

/rpt_family="MER101"

complement(7867..7941)

/rpt_family="Charlie1"

complement(7911..8074)

/rpt_family="Charlie1"

complement(8122..8326)

/rpt_family="L2"

complement(8450..8745)

/rpt_family="AluSg"

complement(9009..9077)

/rpt_family="MIR"

9195..9269

/rpt_family="FLAM"

9276..9422

/rpt_family="AluCo"

9423..9659

/rpt_family="AluDo"

complement(9819..9959)

/rpt_family="L2"

```

repeat_region complement(10121..10231)
repeat_region /rpt_family="MIR"
repeat_region 10740..10923
repeat_region /rpt_family="MER5A"
repeat_region complement(10947..11191)
repeat_region /rpt_family="L2"
repeat_region 11396..11619
repeat_region /rpt_family="MIR"
repeat_region 11646..11775
repeat_region /rpt_family="MIR3"
repeat_region 12444..12474
repeat_region /rpt_family="(CA)n"
repeat_region complement(12619..12834)
repeat_region /rpt_family="MER30"
repeat_region complement(12911..13257)
repeat_region /rpt_family="MER58B"
repeat_region 13529..13839
repeat_region /rpt_family="AluSx"
repeat_region complement(14058..14147)
repeat_region /rpt_family="L2"
repeat_region 14721..15031
repeat_region /rpt_family="AluSp"
repeat_region complement(15717..15794)
repeat_region /rpt_family="MIR"
repeat_region 16155..16459
repeat_region /rpt_family="AluCo"
repeat_region 16463..16762
repeat_region /rpt_family="AluSg"
repeat_region complement(17316..17378)
repeat_region /rpt_family="MIR"
repeat_region 17693..17900
repeat_region /rpt_family="AluCb"
repeat_region 17934..17968
repeat_region /rpt_family="AT-rich"
repeat_region complement(18398..18456)
repeat_region /rpt_family="MIR"
repeat_region 19382..19682
repeat_region /rpt_family="MSTD"
repeat_region 19683..19985
repeat_region /rpt_family="AluSx"
repeat_region 19987..20282
repeat_region /rpt_family="AluSc"
repeat_region 20283..20567
repeat_region /rpt_family="AluY"
repeat_region 20568..20616
repeat_region /rpt_family="MSTD"
repeat_region complement(21371..21672)
repeat_region /rpt_family="AluSx"
repeat_region 22147..22445
repeat_region /rpt_family="AluY"
repeat_region 22493..22790
repeat_region /rpt_family="AluSx"
repeat_region complement(22969..23165)
repeat_region /rpt_family="MIR"
repeat_region 23223..23264
repeat_region /rpt_family="L2"
repeat_region complement(23265..23616)
repeat_region /rpt_family="THE1B"
repeat_region 23617..23696
repeat_region /rpt_family="L2"
repeat_region 25006..25336
repeat_region /rpt_family="AluSx"
repeat_region complement(25722..25905)
repeat_region /rpt_family="AluCo"

```

Query Match

Best Local Similarity 87.6%; Score 18.4; DB 9; Length 136372;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGCGATTGGAGATTGCA 21

DB 58131 AGCGATTGGAAGATTGCA 58112

FEATURES

source

Location/Qualifiers
1. 5518
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 1383 a 142 c 1452 g 2541 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 5518;
Best Local Similarity 95.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAGGGGATTGGAGATTGTG 20
|||||
46 TAGGGGATTGGAGATTGTG 65

RESULT 2 AX344611 5518 bp DNA linear PAT 01-FEB-2002
LOCUS AX344611
DEFINITION Sequence 36 from Patent WO0200927.
ACCESSION AX344611
VERSION AX344611.1 GI:18492497

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with development genes
JOURNAL Patent: WO 0200927-A 36 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source 1. 5518
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 1383 a 142 c 1452 g 2541 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 5518;
Best Local Similarity 95.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAGGGGATTGGAGATTGTG 20
|||||
46 TAGGGGATTGGAGATTGTG 65

RESULT 3 AX345119 5518 bp DNA linear PAT 01-FEB-2002
LOCUS AX345119
DEFINITION Sequence 190 from Patent WO0200928.
ACCESSION AX345119
VERSION AX345119.1 GI:18493005

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequences.

REFERENCE 1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 190 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source 1. 5518
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 1383 a 142 c 1452 g 2541 t
ORIGIN

Query Match 87.6%; Score 18.4; DB 6; Length 5518;

Best Local Similarity 95.0%; Pred. No. 79;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 TAGGGGATTGGAGATTGTG 20
|||||
46 TAGGGGATTGGAGATTGTG 65

RESULT 4 AC103828 136372 bp DNA linear PRI 20-APR-2002
LOCUS AC103828
DEFINITION Homo sapiens chromosome 11, clone CTD1-2183M20, complete sequence.
AC103828
VERSION AC103828.2 GI:20219181

KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens

REFERENCE 1
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Mammals: Eutheria: Primates: Catarrhini: Homiidae: Homo.
JOURNAL 1 (bases 1 to 136372)
REFERENCE 2 (bases 1 to 136372)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone CTD1-2183M20
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 136372)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukhaltier, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagob, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamet, A., Karatas, A., Kells, C., LaRoque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margolis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Melitrim, J.,
Meneus, L., Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schpack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Themann, N., Stojanovic, N.,
Strauss, N., Sudramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

JOURNAL Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 136372)

REFERENCE

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhaltier, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagob, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamet, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margolis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Melitrim, J., Meneus, L.,
Mihova, T., Mlenka, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Plerre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schpack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Themann, N., Stojanovic, N.,
Strauss, N., Sudramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 74.5828 Seconds

(without alignments)
8194.363 Million cell updates/sec

Title: US-09-477-082-31

Perfect score: 21
Sequence: 1 tagggattgagattgtga 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18.4	87.6	5518	6	AX281276	AX281276 Sequence
2	18.4	87.6	5518	6	AX344611	AX344611 Sequence
3	18.4	87.6	5518	6	AX345119	AX345119 Sequence
4	18.4	87.6	135372	10	AC103828	AC103828 Homo sapi
5	18.4	87.6	182686	10	AC121986	AC121986 Mus muscu
6	18.4	87.6	186048	9	AC074190	AC074190 Homo sapi
7	18.4	87.6	188868	9	AC023232	AC023232 Homo sapi
8	18.4	87.6	214787	2	AC123853	AC123853 Mus muscu
9	18.4	85.7	70122	9	AC091938	AC091938 Homo sapi
10	18.4	85.7	179686	2	AC023987	AC023987 Homo sapi
11	17.8	84.8	4369	9	AK096637	AK096637 Homo sapi
12	17.8	84.8	13377	6	AX251229	AX251229 Sequence
13	17.8	84.8	13377	6	AX346564	AX346564 Sequence
14	17.8	84.8	20303	3	AC006749	AC006749 Caenorhab
15	17.8	84.8	115291	2	AL136453	AL136453 Human DNA
16	17.8	84.8	119987	2	CNS08086	CNS08086 Cryza sat
17	17.8	84.8	123701	9	AL139236	AL139236 Human DNA
18	17.8	84.8	146790	2	AC079342	AC079342 Homo sapi
19	17.8	84.8	160650	3	AC010695	AC010695 Drosophi
20	17.8	84.8	160826	9	AC096898	AC096898 Homo sapi
21	17.8	84.8	162402	3	AC010574	AC010574 Drosophi
22	17.8	84.8	165379	2	AC069265	AC069265 Homo sapi
23	17.8	84.8	168506	9	AC007404	AC007404 Homo sapi
24	17.8	84.8	184375	2	AC026219	AC026219 Homo sapi
25	17.8	84.8	190855	2	AL844490	AL844490 Mus muscu
26	17.8	84.8	190964	10	AL671848	AL671848 Mouse DNA
27	17.8	84.8	197121	2	AC025520	AC025520 Mus muscu
28	17.8	84.8	215233	2	AC116590	AC116590 Mus muscu
29	17.8	84.8	224468	2	AL596455	AL596455 Mus muscu
30	17.8	84.8	224790	2	AL683819	AL683819 Mus muscu
31	17.8	84.8	279075	3	AE003530	AE003530 Drosophi
32	17.8	82.9	139885	9	AC009232	AC009232 Homo sapi
33	17.4	82.9	186709	2	AC127994	AC127994 Rattus no
34	17.4	82.9	193277	10	AL596096	AL596096 Mouse DNA
35	17.4	81.0	16914	6	AX251953	AX251953 Sequence
36	17.4	81.0	16914	6	AX348748	AX348748 Sequence
37	17.4	81.0	120310	2	AP004132	AP004132 Oryza sat
38	17.4	81.0	149807	2	AP004998	AP004998 Oryza sat
39	17.4	81.0	171705	9	AC104166	AC104166 Homo sapi
40	17.4	81.0	218652	9	AC096922	AC096922 Homo sapi
41	17.4	80.0	353	11	G50981	G50981 SHGC-80091
42	16.8	80.0	477	8	AY094051	AY094051 Arabidops
43	16.8	80.0	639	8	GA040600	GA040600 Gibberella
44	16.8	80.0	836	8	AY069896	AY069896 Arabidops
45	16.8	80.0	836	8	AY069896	AY069896 Arabidops

ALIGNMENTS

RESULT 1
AX281276
LOCUS
DEFINITION
Sequence 18 from Patent WO0177164.
ACCESSION
AX281276.1 GI:16608531
VERSION
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.

REFERENCE
1
Olek, A., Piepenbrock, C. and Berlin, K.
AUTHORS
TITLE
Diagnosis of diseases associated with apoptosis
Patent: WO 0177164-A 18 OCT-2001;
JOURNAL
Epigenomics AG (DE)

5518 bp DNA linear PAT 02-NOV-2001

Pred. No. is the number of results predicted by chance to have a

US-09-911-904-331

Query Match

Best Local Similarity 69.1%; Score 15.2; DB 9; Length 116;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATCTACATCGAAGC 20

DB 95 CGTATCTACATCGAAGC 114

RESULT 14

US-09-974-300-2342/c
Sequence 2342, Application US/09974300
Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berkas, Randy M.
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2342
LENGTH: 810
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-2342Query Match 69.1%; Score 15.2; DB 10; Length 810;
Best Local Similarity 85.0%; Pred. No. 1.3e+03;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TATATCTACATCGAAGC 22

DB 742 TATGCTACATCGAATGA 723

RESULT 15

US-09-758-269-9/c
Sequence 9, Application US/09758269
Patent No. US20020104120A1

GENERAL INFORMATION:

APPLICANT: IUCHI, SATOSHI
APPLICANT: KOBAYASHI, MASATOMO
APPLICANT: SHINOZAKI, KAZUO
TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN
FILE REFERENCE: 3914-3
CURRENT APPLICATION NUMBER: US/09/758,269
CURRENT FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: JP 2001-003476
PRIOR FILING DATE: 2001-01-11
PRIOR APPLICATION NUMBER: JP 2000-010056
PRIOR FILING DATE: 2000-01-13
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 1734
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1731)
US-09-758-269-9

Query Match 69.1%; Score 15.2; DB 10; Length 1734;

Best Local Similarity 85.0%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GTATATCTACATCGAAGC 21

DB 1471 GTATATCTACCTTCGCAATG 1452

Search completed: July 8, 2003, 06:04:14
Job time : 31.1402 secs

US-09-973-945A-25

Query Match 70.0%; Score 15.4; DB 9; Length 29;
Best Local Similarity 94.1%; Pred. No. 7.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAAA 19
DB 10 TATATCTATATTCGAAA 26

RESULT 10

US-09-973-945A-26/C
Sequence 26, Application US/09973945A
Patent No. US20020169297A1
GENERAL INFORMATION:
APPLICANT: University of Kentucky Research Foundation
APPLICANT: XIE, Shuheng
APPLICANT: XIE, Mingtang
TITLE OF INVENTION: Genetic Insulator for Preventing Influence By Another Gene
FILE REFERENCE: 050229-0287
CURRENT APPLICATION NUMBER: US/09/973,945A
CURRENT FILING DATE: 2001-10-11
PRIOR APPLICATION NUMBER: 60/241,735
PRIOR FILING DATE: 2000-10-20
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn version 3.1
SEQ ID NO 26
LENGTH: 29
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-973-945A-26

Query Match 70.0%; Score 15.4; DB 9; Length 29;
Best Local Similarity 94.1%; Pred. No. 7.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAAA 19
DB 20 TATATCTATATTCGAAA 4

RESULT 11

US-08-781-986A-1583/C
Sequence 1583, Application US/08781986A
Publication No. US20030054436A1
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
NUMBER OF SEQUENCES: 5255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248BP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 1583:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-1583

Query Match 70.0%; Score 15.4; DB 7; Length 348;
Best Local Similarity 94.1%; Pred. No. 9.6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAAA 19
DB 206 TATATCTACATTAGAAA 190

RESULT 12

US-09-911-904-216
Sequence 216, Application US/09911904
Publication No. US20030096234A1
GENERAL INFORMATION:
APPLICANT: Fair, Spencer B.
APPLICANT: Pickett, Gavin G.
APPLICANT: Neft, Robin Eileen
APPLICANT: Dunn, II, Robert Thomas
TITLE OF INVENTION: CANINE TOXICITY GENES
FILE REFERENCE: 400742000200
CURRENT APPLICATION NUMBER: US/09/911,904
CURRENT FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: US 60/220,057
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 216
LENGTH: 116
TYPE: DNA
ORGANISM: Canis familiaris
US-09-911-904-216

Query Match 69.1%; Score 15.2; DB 9; Length 116;
Best Local Similarity 85.0%; Pred. No. 1e+03;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATATCTACATTCGAAC 20
DB 95 CGTATGCTACAGTCGAAGC 114

RESULT 13

US-09-911-904-331
Sequence 331, Application US/09911904
Publication No. US20030096234A1
GENERAL INFORMATION:
APPLICANT: Fair, Spencer B.
APPLICANT: Pickett, Gavin G.
APPLICANT: Neft, Robin Eileen
APPLICANT: Dunn, II, Robert Thomas
TITLE OF INVENTION: CANINE TOXICITY GENES
FILE REFERENCE: 400742000200
CURRENT APPLICATION NUMBER: US/09/911,904
CURRENT FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: US 60/220,057
PRIOR FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 331
LENGTH: 116
TYPE: DNA
ORGANISM: Canis familiaris

```

US-10-239-676-162
Query Match          70.9%: Score 15.6; DB 9; Length 15305;
Best Local Similarity 81.8%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 CGTATCTACATTGCAACGA 22
||| ||||||| |||||
Db      8794 CGAACACTCACCATTGAAACGA 8773

RESULT 8
US-09-738-626-1
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKABARA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENO, AKIHITO
APPLICANT: IKEDA, MASATO
APPLICANT: OKAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match          70.9%: Score 15.6; DB 9; Length 3309400;
Best Local Similarity 81.8%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 CGTATCTACATTGCAACGA 22
||| ||||||| |||||
Db      2400809 CGCATAGCCACATTGCAAAATGA 2400830

RESULT 9
US-09-973-945A-25
Sequence 25, Application US/09973945A
Patent No. US20020169297A1
GENERAL INFORMATION:
APPLICANT: University of Kentucky Research Foundation
APPLICANT: XIE, Shusheng
APPLICANT: XIE, Mingtang
TITLE OF INVENTION: Genetic Insulator for Preventing Influence By Another Gene
FILE REFERENCE: 050229-0287
CURRENT APPLICATION NUMBER: US/09/973,945A
CURRENT FILING DATE: 2001-10-11
PRIOR APPLICATION NUMBER: 60/241,735
PRIOR FILING DATE: 2000-10-20
NUMBER OF SEQ ID NOS: 37
SOFTWARE: PatentIn version 3.1
SEQ ID NO 25
LENGTH: 29
TYPE: DNA
ORGANISM: Arabidopsis thaliana

```

APPLICANT: Hu, Weng
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroskin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 168
LENGTH: 2909
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-168

Query Match 73.6%; Score 16.2; DB 9; Length 2909;
Best Local Similarity 85.7%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATATCTACATTCGAACG 21
DB 2061 CTTATGCTACATTCGAACG 2041

RESULT 3
US-10-128-714-5168/c
Sequence 5168, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Weng
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroskin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31
NUMBER OF SEQ ID NOS: 8603
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5168
LENGTH: 3224
TYPE: DNA
ORGANISM: Aspergillus fumigatus
US-10-128-714-5168

Query Match 73.6%; Score 16.2; DB 9; Length 3224;
Best Local Similarity 85.7%; Pred. No. 5.4e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CGTATATCTACATTCGAACG 21
DB 2376 CTTATGCTACATTCGAACG 2356

RESULT 4
US-10-108-605-70/c
Sequence 70, Application US/10108605
Patent No. US20020160934A1
GENERAL INFORMATION:
APPLICANT: Broadus, Julie
APPLICANT: Stam, Lynn
APPLICANT: Bachmann, Jane
APPLICANT: Kamdar, Kim
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD
FILE REFERENCE: 3113B
CURRENT APPLICATION NUMBER: US/10/108,605
CURRENT FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: US 09/761,142
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/176,418
PRIOR FILING DATE: 2000-01-14
NUMBER OF SEQ ID NOS: 361
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 11527
TYPE: DNA
ORGANISM: Drosophila melanogaster
US-10-108-605-70

Query Match 73.6%; Score 16.2; DB 9; Length 11527;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATATCTACATTCGAACG 21
DB 6421 CGTTATCTCTACATTCGAATG 6401

RESULT 5
US-09-974-300-1167
Sequence 1167, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1167
LENGTH: 1356
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1167

Query Match 70.9%; Score 15.6; DB 10; Length 1356;
Best Local Similarity 81.8%; Pred. No. 9.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CGTATATCTACATTCGAACG 22
DB 685 CGTATCTCTACATTCGAACG 706

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 11.1402 seconds
(without alignments)
3082.690 Million cell updates/sec

Title: US-09-477-082-30

Perfect score: 22
Sequence: 1 cgtatctcattcgaaacga 22

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	17.2	78.2	2919	US-09-938-842A-2471	Sequence 2471, Ap
2	16.2	73.6	2909	US-10-128-714-168	Sequence 168, App
3	16.2	73.6	3224	US-10-128-714-5168	Sequence 5168, Ap
4	16.2	73.6	11527	US-10-108-605-70	Sequence 70, Appl
5	15.6	70.9	1356	US-09-974-300-1167	Sequence 1167, Ap
6	15.6	70.9	1899	US-09-738-626-2486	Sequence 2486, Ap
7	15.6	70.9	15306	US-10-239-676-162	Sequence 162, App
8	15.6	70.9	3309400	US-09-738-626-1	Sequence 1, Appl
9	15.4	70.0	29	US-09-973-945A-25	Sequence 25, Appl
10	15.4	70.0	29	US-09-973-945A-26	Sequence 26, Appl
11	15.4	70.0	348	US-08-781-986A-1583	Sequence 1583, Ap
12	15.2	69.1	116	US-09-911-904-216	Sequence 216, App
13	15.2	69.1	116	US-09-911-904-331	Sequence 331, App
14	15.2	69.1	810	US-09-974-300-2342	Sequence 2342, Ap
15	15.2	69.1	1734	US-09-758-269-9	Sequence 9, Appl
16	15.2	69.1	1935	US-10-097-065-117	Sequence 117, App
17	15.2	69.1	3718	US-10-128-714-5267	Sequence 5267, Ap
18	15.2	69.1	6071	US-10-239-676-78	Sequence 78, Appl
19	15.2	69.1	8333	US-10-239-676-114	Sequence 114, App

20	15.2	69.1	31412	US-10-109-551-3	Sequence 3, Appl
21	15.2	69.1	659158	US-09-771-208-20	Sequence 20, Appl
22	14.8	67.3	199	US-09-747-377-412	Sequence 412, App
23	14.8	67.3	199	US-10-105-613-412	Sequence 412, App
24	14.8	67.3	234	US-09-974-300-4075	Sequence 4075, Ap
25	14.8	67.3	1064	US-09-938-842A-4152	Sequence 4152, Ap
26	14.8	67.3	1140	US-09-893-518A-116	Sequence 116, App
27	14.8	67.3	1459	US-10-198-846-13846	Sequence 13846, A
28	14.8	67.3	1992	US-10-121-032-57	Sequence 57, Appl
29	14.8	67.3	1992	US-10-093-037-57	Sequence 57, Appl
30	14.8	67.3	2043	US-10-121-032-11	Sequence 11, Appl
31	14.8	67.3	2043	US-10-093-037-11	Sequence 11, Appl
32	14.8	67.3	2097	US-09-801-368-373	Sequence 373, App
33	14.8	67.3	2237	US-09-832-292-11	Sequence 11, Appl
34	14.8	67.3	2237	US-09-994-485-7	Sequence 7, Appl
35	14.8	67.3	3237	US-10-071-766-39	Sequence 39, Appl
36	14.8	67.3	3595	US-09-089-818B-6	Sequence 6, Appl
37	14.8	67.3	6158	US-10-239-676-23	Sequence 23, Appl
38	14.8	67.3	7004	US-10-239-676-120	Sequence 120, App
39	14.8	67.3	17787	US-09-764-868-1285	Sequence 1285, App
40	14.8	67.3	17787	US-09-764-891-5648	Sequence 5648, Ap
41	14.8	67.3	17787	US-09-764-891-6051	Sequence 6051, Ap
42	14.8	67.3	22299	US-10-091-577-469	Sequence 469, App
43	14.8	67.3	22299	US-09-764-891-6605	Sequence 6605, App
44	14.8	67.3	28313	US-09-764-877-3194	Sequence 3194, App
45	14.6	66.4	324	US-09-803-719-19	Sequence 19, Appl

ALIGNMENTS

```

RESULT 1
US-09-938-842A-2471
; Sequence 2471, Application US/09938842A
; Patent No. US20020160378A1
;
GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Krep, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
;
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;
FILE REFERENCE: SCRIPT300-3
;
CURRENT APPLICATION NUMBER: US/09/938,842A
;
PRIOR FILING DATE: 2001-08-24
;
PRIOR APPLICATION NUMBER: US 60/227,866
;
PRIOR FILING DATE: 2000-08-24
;
PRIOR APPLICATION NUMBER: US 60/264,647
;
PRIOR FILING DATE: 2001-01-16
;
PRIOR APPLICATION NUMBER: US 60/300,111
;
PRIOR FILING DATE: 2001-06-22
;
NUMBER OF SEQ ID NOS: 5379
;
SEQ ID NO 2471
;
LENGTH: 2919
;
TYPE: DNA
;
ORGANISM: Arabidopsis thaliana
;
US-09-938-842A-2471
;
Query Match          78.2%; Score 17.2; DB 9; Length 2919;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
QY      1 CGTATCTGACATTCGAACGA 22
;      1 ||||| ||||| |||||
Db      1918 CTTATCTGCATTCGAACGA 1939
;
RESULT 2
US-10-128-714-168/c
; Sequence 168, Application US/10128714
; Patent No. US20030119013A1
;
GENERAL INFORMATION:
; APPLICANT: Jiang, Bo

```

QY 1 CGTATCTACATTCGAACG 21
 || ||||| ||||| |||||
 Db 427 CGCATATCACATTCGAATG 447

RESULT 15
 US-09-605-785-633/C
 ; Sequence 633, Application US/09605785
 ; Patent No. 6321716
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yugu
 ; APPLICANT: Henderson, Robert A.
 ; APPLICANT: Kalos, Michael D.
 ; APPLICANT: Fanger, Gary R.
 ; APPLICANT: Retter, Marc W.
 ; APPLICANT: Stolk, John A.
 ; APPLICANT: Day, Craig H.
 ; APPLICANT: Vedwick, Thomas S.
 ; APPLICANT: Carter, Darrick
 ; APPLICANT: Li, Samuel
 ; APPLICANT: Wang, Aljun
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Hepler, William
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
 ; FILE REFERENCE: 210121.427C16
 ; CURRENT APPLICATION NUMBER: US/09/605,785
 ; CURRENT FILING DATE: 2000-06-27
 ; NUMBER OF SEQ ID NOS: 835
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 633
 ; LENGTH: 630
 ; TYPE: DNA
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(630)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-09-605-785-633

Query Match 66.4%; Score 14.6; DB 4; Length 630;
 Best local Similarity 81.0%; Pred. No. 1.6e+02;
 Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GTATATCTACATTCGAACG 22
 ||||| ||||| ||||| |||||
 Db 366 GTATATCTACATTCGATG 346

Search completed: July 8, 2003, 05:56:44
 Job time : 27.2769 secs

APPLICATION NUMBER: US/08/468,347
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 469 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-468-347-23

Query Match
Best Local Similarity 66.4%; Score 14.6; DB 1; Length 469;
Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 CGTATCTACTTCGAAACG 21
427 CGCATATCAACATTCAAATG 447

RESULT 10
US-08-226-264-25
Sequence 25, Application US/08226264
Patent No. 5801017
GENERAL INFORMATION:
APPLICANT: Weiber, Moshe M.
APPLICANT: Zeelon, Elisha P.
APPLICANT: Levanon, Avigdor
APPLICANT: Guy, Rachel
APPLICANT: Goldlust, Arie
APPLICANT: Rigbi, Meir
APPLICANT: Panet, Amos
APPLICANT: Fischer, Meir
TITLE OF INVENTION: PRODUCTION OF RECOMBINANT FACTOR XA
TITLE OF INVENTION: INHIBITORS OF LEECH HIRUDO MEDICINALIS
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/226,264
FILING DATE: 08-APR-94
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 40017-A/JPW/GJG
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525

TELEX:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 469 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-226-264-25

Query Match
Best Local Similarity 66.4%; Score 14.6; DB 1; Length 469;
Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 CGTATCTACTTCGAAACG 21
427 CGCATATCAACATTCAAATG 447

RESULT 11
US-08-467-389-23
Sequence 23, Application US/08467389
Patent No. 5824641
GENERAL INFORMATION:
APPLICANT: Zeelon, Elisha P.
APPLICANT: Weiber, Moshe M.
APPLICANT: Levanon, Avigdor
APPLICANT: Guy, Rachel
APPLICANT: Goldlust, Arie
APPLICANT: Rigbi, Meir
APPLICANT: Panet, Amos
APPLICANT: Fischer, Meir
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR XA
TITLE OF INVENTION: INHIBITORY ACTIVITY
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,389
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/225,442
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0317/43020-A/JPW/EAB
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-977-9550
TELEFAX: 212-664-0525
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 469 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-467-389-23

Query Match
Best Local Similarity 66.4%; Score 14.6; DB 1; Length 469;
Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Patent No. 6030832
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Bartel, Paul L.
APPLICANT: Teng, David H.-F.
APPLICANT: Tavligian, Sean V.
TITLE OF INVENTION: A Carboxy-Terminal BRCA1 Interacting
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701 East
CITY: Washington
STATE: DC
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,703
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Stephen A.
REGISTRATION NUMBER: 38,609
REFERENCE/DOCKET NUMBER: 2318-0174
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-624-1589
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2691
US-08-975-703-5

Query Match 67.3%; Score 14.8; DB 3; Length 2694;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSTATCTACATCGAAA 19
DB 1103 GSTATCTAGATTGAAA 1120

RESULT 8
US-09-515-884-5
Sequence 5, Application US/09515884
Patent No. 6235263
GENERAL INFORMATION:
APPLICANT: Mong, Alexander K.C.
APPLICANT: Bartel, Paul L.
APPLICANT: Teng, David H.-F.
APPLICANT: Tavligian, Sean V.
TITLE OF INVENTION: A Carboxy-Terminal BRCA1 Interacting
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701 East

CITY: Washington
STATE: DC
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/515,884
FILING DATE: 29-Feb-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/975,703
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Stephen A.
REGISTRATION NUMBER: 38,609
REFERENCE/DOCKET NUMBER: 2318-0174
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-624-1589
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2691
US-09-515-884-5

Query Match 67.3%; Score 14.8; DB 4; Length 2694;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GSTATCTACATCGAAA 19
DB 1103 GSTATCTAGATTGAAA 1120

RESULT 9
US-08-468-347-23
Sequence 23, Application US/08468347
Patent No. 5783421
GENERAL INFORMATION:
APPLICANT: Zeeion, Elisha P.
APPLICANT: Werber, Moshe M.
APPLICANT: Levanon, Avigdor
TITLE OF INVENTION: NOVEL POLYPEPTIDE HAVING FACTOR Xa
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1034:
SEQUENCE CHARACTERISTICS:
LENGTH: 2018 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...2018
US-09-221-017B-1034

Query Match 67.3%; Score 14.8; DB 4; Length 2018;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TATCTACATTCGAACGA 22
|||||
Db 1169 TATCGACATTCGATACGA 1152

RESULT 5
US-09-134-078-11
Sequence 11, Application US/09134078
Patent No. 6368844
GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 2043 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA

FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...2040
US-09-134-078-11

Query Match 67.3%; Score 14.8; DB 4; Length 2043;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATATCTACATTCGAACG 21
|||||
Db 1754 ACATCTACATTCGACAA 1771

RESULT 6
US-08-914-999-7/C
Sequence 7, Application US/08914999
Patent No. 6346406
GENERAL INFORMATION:
APPLICANT: Ryzanov, Alexey G.
APPLICANT: Pavur, Karen S.
TITLE OF INVENTION: ELONGATION FACTOR-2 KINASE (EF-2 KINASE)
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/914,999
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 601-1-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2237 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Dictyostelium discoideum
US-08-914-999-7

Query Match 67.3%; Score 14.8; DB 4; Length 2237;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATATCTACATTCGAA 19
|||||
Db 1745 GTATATTCATTCGCAA 1728

RESULT 7
US-08-975-703-5
Sequence 5, Application US/08975703

```
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 4212 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMYXAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...4212
US-09-221-017B-39

Query Match
Best Local Similarity 69.1%; Score 15.2; DB 4; Length 4212;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22
DB 2978 TATATATAATTAGAACGA 2959

RESULT 3
US-09-134-078-57
Sequence 57, Application US/09134078
Patent No. 6368844
GENERAL INFORMATION:
APPLICANT: Bylina, Edward J.
TITLE OF INVENTION: GLYCOSIDASE ENZYMES
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gray Cary Ware & Freidenrich LLP
STREET: 4365 Executive Drive, Suite 1600
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,078
FILING DATE: 13-AUG-1998
```

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/349,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 1992 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: Genomic DNA
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...1998
US-09-134-078-57

Query Match
Best Local Similarity 67.3%; Score 14.8; DB 4; Length 1992;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATATCTACATTCGAACG 21
DB 1703 ACATCTACATTCGAACG 1720

RESULT 4
US-09-221-017B-1034/C
Sequence 1034, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 5.27694 Seconds
(without alignments)
1278.561 Million cell updates/sec

Title: US-09-477-082-30

Perfect score: 22

Sequence: 1 cgtatctacattcgaacga 22

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCRTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfilest.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.2	69.1	999	4	US-09-134-001C-1178
C 2	15.2	69.1	4212	4	US-09-221-017B-39
C 3	14.8	67.3	1992	4	US-09-134-078-57
C 4	14.8	67.3	2018	4	US-09-221-017B-1034
C 5	14.8	67.3	2043	4	US-09-134-078-11
C 6	14.8	67.3	2237	4	US-08-914-999-7
C 7	14.8	67.3	2694	3	US-08-975-703-5
C 8	14.8	67.3	2694	3	US-08-915-884-5
C 9	14.6	66.4	469	1	US-08-468-347-23
C 10	14.6	66.4	469	1	US-08-226-264-25
C 11	14.6	66.4	469	1	US-08-467-389-23
C 12	14.6	66.4	469	2	US-08-779-378-23
C 13	14.6	66.4	469	2	US-08-469-219-23
C 14	14.6	66.4	469	2	US-09-228-152-23
C 15	14.6	66.4	630	4	US-09-605-785-633
C 16	14.6	66.4	780	4	US-09-134-001C-1631
C 17	14.6	66.4	1410	2	US-08-975-316-86
C 18	14.6	66.4	1410	2	US-09-615-192A-86
C 19	14.6	66.4	2112	4	US-09-232-160-12
C 20	14.6	66.4	2728	4	US-09-232-160-1
C 21	14.6	66.4	3504	4	US-08-857-076-47
C 22	14.6	66.4	3959	1	US-08-474-067-1
C 23	14.6	66.4	3959	2	US-08-474-068A-1
C 24	14.6	66.4	3959	2	US-08-472-481-1
C 25	14.6	66.4	11811	4	US-09-078-294-7
C 26	14.6	65.5	369	1	US-09-053-945-3
C 27	14.4	65.5	607	1	US-09-053-945-1

C 28	14.4	65.5	607	1	US-08-061-314A-1	Sequence 1, Appl1
C 29	14.4	65.5	711	4	US-09-134-001C-1834	Sequence 1834, Ap
C 30	14.4	65.5	2661	4	US-09-134-001C-1659	Sequence 1659, Ap
C 31	14.4	65.5	5055	4	US-09-242-632A-13	Sequence 13, Appl
C 32	14.4	65.5	31491	4	US-09-360-186-1	Sequence 11, Appl1
C 33	14.2	64.5	213	4	US-09-134-001C-1113	Sequence 1113, Ap
C 34	14.2	64.5	222	4	US-09-134-001C-1599	Sequence 599, App
C 35	14.2	64.5	470	4	US-09-615-192A-189	Sequence 189, App
C 36	14.2	64.5	609	1	US-08-268-072-2	Sequence 2, Appl1
C 37	14.2	64.5	609	1	US-08-690-721-2	Sequence 27, Appl
C 38	14.2	64.5	684	1	US-08-226-264-27	Sequence 878, App
C 39	14.2	64.5	836	4	US-09-221-017B-878	Sequence 581, App
C 40	14.2	64.5	1314	4	US-09-134-001C-581	Sequence 470, App
C 41	14.2	64.5	1521	4	US-09-134-001C-470	Sequence 1, Appl1
C 42	14.2	64.5	1812	2	US-08-669-536-1	Sequence 1, Appl1
C 43	14.2	64.5	2000	1	US-09-041-075A-1	Sequence 3, Appl1
C 44	14.2	64.5	3030	4	US-09-693-146-3	Sequence 1, Appl1
C 45	14.2	64.5	3927	4	US-09-293-238B-1	

ALIGNMENTS

RESULT 1
US-09-134-001C-1178/c
Sequence 1178, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134, 001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064, 964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055, 779
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1178
LENGTH: 999
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1178

Query Match 69.1%; Score 15.2; DB 4; Length 999;
Best Local Similarity 85.0%; Pred. No. 87;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Caps 0;

Cy 3 TATATCTACATTCGAACGA 22
Db 508 TATATCTACATTCGAACGA 489

RESULT 2
US-09-221-017B-39/c
Sequence 39, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows

TITLE
P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://comgenomics.ucdavis.edu/
unpublished (2002)

JOURNAL
COMMENT
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QG16 row: M column: 15.
Location/Qualifiers

FEATURES

1. 599
/organism="Lactuca sativa"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QG16M15"
/lab_host="E.coli"
/lab_lib="QG-ABCDI lettuce salinas"
/note="Vector: pBRCDNA5flab: The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG_LIB-QG-ABCDI lettuce salinas
TAG_TISSUE=roots
TAG_SEQ=GTTCACGCGC"

BASE COUNT 201 a 113 c 121 g 164 t

ORIGIN

Query Match 76.4%; Score 16.8; DB 14; Length 599;
Best Local Similarity 90.0%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATTGGAACGA 22
||||| ||||||| |||
Db 440 TATATCTACATTGGAACGA 459

RESULT 15

LOCUS AQ311480 639 bp DNA linear GSS 22-DEC-1998
DEFINITION CITR1-EL-2526L19.TR CITR1-EL Homo sapiens genomic clone 2526L19,
DNA sequence.

ACCESSION AQ311480
VERSION AQ311480.1 GI:4039833
KEYWORDS GSS.

SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 639)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and
Venter, J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL unpublished (1998)
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:
http://www.tigr.org/tcb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES

Location/Qualifiers
1. 639
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2526L19"
/clone_lib="CITR1-EL"
/sex="male"
/cell_type="sperm"
/note="Vector: pBeloBAC11, Site_1: EcoRI, Site_2: EcoRI;
Caltech Human BAC Library D"

BASE COUNT 205 a 122 c 142 g 170 t

ORIGIN

Query Match 76.4%; Score 16.8; DB 17; Length 639;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATTGGAACGA 22
||||| ||||||| |||
Db 206 TATATCTATATTCGAACACTA 225

Search completed: July 8, 2003, 05:51:22
Job time: 170.879 secs

FEATURES Seq primer: M13uni primer for 3'end.
Location/Qualifiers

1..324
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone_1lb="HM04613u"
/clone_1lb="Hordeum vulgare Barke roots"
/tissue-type="roots"
/lab_host="XLOLR"
/note="Vector: plasmid pK-CMV; Site_1: EcoRI; Site_2: XhoI; mRNA was made from roots of spring barley variety 'Barke', a high quality malting variety. Roots were grown for two days on filter paper at room temperature. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA).
NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb. Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'- and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp"

BASE COUNT 73 a 89 c 49 g 108 t 5 others

Query Match 76.4%; Score 16.8; DB 9; Length 324;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 CGTATCTACATTCGAAC 20
Db 157 CGTATCTACATTCGAAC 176

RESULT 12
LOCUS AU037723 451 bp mRNA linear EST 29-MAR-1999

DEFINITION AU037723 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoidium cDNA clone SSE216, mRNA sequence.

ACCESSION AU037723
VERSION AU037723
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 451)
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

TITLE
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development

JOURNAL
MEDLINE
COMMENT
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = "Dictyostelium discoideum cDNA project in Japan".

FEATURES
Source
1..451
Location/Qualifiers
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSE216"
/clone_1lb="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 170 a 44 c 82 g 155 t

Query Match 76.4%; Score 16.8; DB 9; Length 451;

Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TATATCTACATTCGAACGA 22
Db 336 TATATCTACATTCGAACGA 354

RESULT 13
LOCUS AU037770 452 bp mRNA linear EST 29-MAR-1999

DEFINITION AU037770 Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoidium cDNA clone SSE272, mRNA sequence.

ACCESSION AU037770
VERSION AU037770.1 GI:3984523
EST.

REFERENCE
AUTHORS
1 (bases 1 to 452)
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Sato, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

TITLE
The Dictyostelium developmental cDNA project: generation and
analysis of expressed sequence tags from the first-finger stage of
development

JOURNAL
MEDLINE
COMMENT
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp
PROJECT = "Dictyostelium discoideum cDNA project in Japan".

FEATURES
source
1..452
Location/Qualifiers
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SSE272"
/clone_1lb="Dictyostelium discoideum SS (H.Urushihara)"
/dev_stage="slug"

BASE COUNT 170 a 45 c 82 g 155 t

Query Match 76.4%; Score 16.8; DB 9; Length 452;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TATATCTACATTCGAACGA 22
Db 336 TATATCTACATTCGAACGA 355

RESULT 14
LOCUS B0874972 599 bp mRNA linear EST 15-AUG-2002

DEFINITION OG1M15.yg.abi OG ABCDI lettuce salinas lactuca sativa cDNA clone
OG1M15, mRNA sequence.

ACCESSION B0874972
VERSION B0874972.1 GI:22261532
EST.

REFERENCE
AUTHORS
1 (bases 1 to 599)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison

FEATURES					
source					
Location/Qualifiers					
1..204					
/organism="Mus musculus"					
/db_xref="taxon:10090"					
/clone="B23031A20"					
/clone_lib="RIKEN full-length enriched, adult male corpora quadrigemina"					
/sex="male"					
/tissue_type="corpora quadrigemina"					
/dev_stage="adult"					
/lab_host="DH108"					
/note="Site_1: Sali; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'					
GAGAGACGACGATCCAAAGCTCTTTTCTTTTCTTTTNVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGACGATTCGCGATTAAATAATTAATCCCCCCCCCCC 3'] cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from LambdaFLC I."					
BASE COUNT					
ORIGIN					
55 a 35 c 28 g 86 t					
Query Match					
Best Local Similarity 90.0%; Pred. No. 3.2e+02;					
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0.					
QY					
2 GGTATCTCATCTCGAAGC 21 Db 56 GTTACTACTGATTCGAATG 37					
RESULT 11					
AL501502 324 bp mRNA linear EST 04-JAN-2001					
LOCUS AL501502 Hordeum vulgare Barke roots Hordeum vulgare cDNA clone HM04G13m 3', mRNA sequence.					
ACCESSION AL501502					
VERSION AL501502.1 GI:12027717					
KEYWORDS					
SOURCE					
ORGANISM					
Hordeum vulgare.					
Hordeum vulgare					
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.					
1 (bases 1 to 324)					
REFERENCE					
AUTHORS Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.					
TITLE EST sequencing and analysis in barley					
JOURNAL Unpublished (2000)					
COMMENT Institute for Plant Genetics and Crop Plant Research Corrensstr. 3, D-06466 Gatersleben, Germany					
Email: michalek@ipk-gatersleben.de, http://pgpc.ipk-gatersleben.de					

COMMENT

Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.

PRIMERS

Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI

FEATURES

Source

Location/Qualifiers
1..1970

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-099107.R"
/sex="male"

/cell_type="lymphoblast"

BASE COUNT 364 a 405 c 473 g 576 t 132 others

ORIGIN

Query Match 80.9%; Score 17.8; DB 17; Length 1970;
Best Local Similarity 86.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

1 CGTATCTACATCGAAGCA 22

Db

1746 CATATCTACATCGAAGCA 1725

RESULT 4

LOCUS

BH582702 791 bp DNA linear GSS 15-DEC-2001

DEFINITION BOGTC24TR BOGT Brassica oleracea genomic clone BOGTC24, DNA

sequence.

ACCESSION

BH582702

VERSION

BH582702.1 GI:17835159

KEYWORDS

GSS

SOURCE

Brassica oleracea.

ORGANISM

Brassica oleracea;
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE

1 (bases 1 to 791)
Town, C.D., Van Aken, S., Utterback, T. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)

AUTHORS

Whole genome shotgun sequencing of Brassica oleracea

TITLE

Whole genome shotgun sequencing of Brassica oleracea

JOURNAL

Unpublished (2001)

COMMENT

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208

Email:

cctown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer:

TR

Class:

sheared ends.

FEATURES

Location/Qualifiers
1..791

Source

/organism="Brassica oleracea"
/strain="T01000D3"
/db_xref="taxon:3712"
/clone="BOGTC24"
/clone_id="BOGT"

BASE COUNT

226 a 209 c 167 g 189 t

ORIGIN

Query Match 79.1%; Score 17.4; DB 17; Length 791;
Best Local Similarity 94.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

4 ATATCTACATCGAAGCA 22

Db

71 ATATCTACATCGAAGCA 89

RESULT 5

LOCUS

B96258 203 bp DNA linear GSS 31-MAR-1998

DEFINITION T23P5TRB TAMU Arabidopsis thaliana genomic clone T23P5, DNA

sequence.

ACCESSION

B96258

VERSION

B96258.1 GI:2998734

KEYWORDS

GSS

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.

REFERENCE

1 (bases 1 to 203)
Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Linher, K., Golden, K.,
Berry, K., Granger, D., Sun, E., Wilde, C., Adams, M.D. and Venter

AUTHORS

, J.C.

TITLE

A BAC End Sequence Database for Identifying Minimal Overlaps in
Arabidopsis Genomic Sequencing. Update 3

JOURNAL

Unpublished (1997)

COMMENT

Other GSSs: T23P5TRB

Contact:

Steve Rounsley

Department of

Eukaryotic Genomics

The Institute for

Genomic Research

9712 Medical Center Dr.,

Rockville, MD 20850, USA

Tel:

301 838 0200

Fax:

301 838 0208

Email:

ronsley@tigr.org

Seq primer:

M13 Reverse

Class:

BAC ends

High quality

sequence stop: 203.

FEATURES

Location/Qualifiers
1..203

Source

/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T23P5"
/clone_id="TAMU"

/sex="hermaphrodite"

/note="Vector: BelobACII; Site_1: HindIII; Site_2: HindIII
; Produced by Rod Wing"

BASE COUNT

71 a 42 c 30 g 60 t

ORIGIN

Query Match

78.2%; Score 17.2; DB 17; Length 203;

Best Local

Similarity 86.4%; Pred. No. 2e+02;

Matches

19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY

1 CGTATCTACATCGAAGCA 22

Db

61 CTAATCTACATCGAAGCA 82

RESULT 6

CNS00V67/c

LOCUS

CNS00V67 660 bp DNA linear GSS 28-JUN-1999

DEFINITION

Arabidopsis thaliana genome survey sequence SP6 end of BAC T94 of
TAMU library from strain Columbia of Arabidopsis thaliana, genomic
survey sequence.

ACCESSION

AI092245

VERSION

AI092245.1 GI:5293399

KEYWORDS

GSS

SOURCE

Arabidopsis thaliana.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.

REFERENCE

1 (bases 1 to 660)

AUTHORS

Salanoubat, M., Choisme, N., Artiguenave, F., Brotier, P., Wincker, P.,

TITLE	RIKEN Mouse ESTs (Konno, H., et
JOURNAL	Unpublished (2000)
COMMENT	Contact: yoshihide Hayashizaki

XX 12-JUL-2002 (first entry)
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 30313.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP10074.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX PT amplicons from chemically treated DNA -
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX AB013410-AB054121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX
XX Sequence 554 BP; 99 A; 57 C; 168 G; 230 T; 0 other;
XX
XX Query Match 73.6%; Score 16.2; DB 24; Length 554;
XX Best Local Similarity 85.7%; Pred. No. 1.3e+02;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 CGTATATCTACATTCGAACG 21
XX DB 109 CGTAAATATTAATTCGAACG 89

RESULT 13
ABQ43723
ID ABQ43723 standard; DNA; 554 BP.
XX
XX ABQ43723;
XX
XX 12-JUL-2002 (first entry)
XX
XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30314.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.
XX
XX Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EP10074.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX PT amplicons from chemically treated DNA -
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridised to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridisation to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridised to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX AB013410-AB054121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.
XX
XX Sequence 554 BP; 230 A; 168 C; 57 G; 99 T; 0 other;
XX
XX Query Match 73.6%; Score 16.2; DB 24; Length 554;
XX Best Local Similarity 85.7%; Pred. No. 1.3e+02;
XX Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 1 CGTATATCTACATTCGAACG 21
XX DB 446 CGTAAATATTAATTCGAACG 466

RESULT 14
ABQ38892
ID ABQ38892 standard; DNA; 598 BP.
XX
XX ABQ38892;
XX
XX 12-JUL-2002 (first entry)
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 25483.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX
XX

RESULT 10
ABQ48138/c
ID ABQ48138 standard; DNA; 513 BP.
XX
XX ABQ48138;
AC
XX
XX 12-JUL-2002 (first entry)
DT
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 34729.
DE
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
PN
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP10074.
PF
XX 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI
XX WPI; 2002-371829/40.
DR
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
PS
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
XX
SQ Sequence 513 BP; 115 A; 41 C; 167 G; 189 T; 1 other;
Query Match 74.5%; Score 16.4; DB 24; Length 513;
Best Local Similarity 94.4%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATATCTACATCGAAGC 21
IIIIIIIIIIIIIIIIIIII
DB 443 ATATCTACATCGAAGC 426

RESULT 11
ABQ48139

ID ABQ48139 standard; DNA; 513 BP.
XX
XX ABQ48139;
AC
XX
XX 12-JUL-2002 (first entry)
DT
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 34730.
DE
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX WO200218632-A2.
PN
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP10074.
PF
XX 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI
XX WPI; 2002-371829/40.
DR
XX
XX Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.
PS
XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
XX
SQ Sequence 513 BP; 189 A; 167 C; 41 G; 115 T; 1 other;
Query Match 74.5%; Score 16.4; DB 24; Length 513;
Best Local Similarity 94.4%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATATCTACATCGAAGC 21
IIIIIIIIIIIIIIIIIIII
DB 71 ATATCTACATCGAAGC 88

RESULT 12
ABQ43722/c
ID ABQ43722 standard; DNA; 554 BP.
XX
XX ABQ43722;
AC

	Matches	18, Conservative	0; Mismatches	2; Indels	0; Gaps
QY	3	TATATCTACATTCGAACGA	22		
Db	496	TATAACTACATCCGAACGA	515		

```

Db      329  TATATATAAATTCGAAACGA 310
          ||||| | | ||||| |||||
RESULT 9

```

RESULT	8
ABQ29578/c	
ID	ABQ29578 standard; DNA; 1447 BP.
XX	
AC	ABQ29578;
XX	
DT	12-JUL-2002 (first entry)
DE	Oligonucleotide for detecting cytosine methylation SEQ ID NO 16159.

XX	ABQ29579;
AC	
XX	12-JUL-2002 (first entry)
DT	
XX	Oligonucleotide for detecting cytosine methylation SEQ ID NO 16170.
DE	
XX	
XX	Human: cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW	drug; side effect; cancer; central nervous system; cardiovascular;

KW Human, cytosine methylation; 5'-cpg-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism
 KW SNP; cell differentiation; ds.

KW SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX

OS	Homo sapiens.
XX	
PN	W0200218632-A2

XX
PD
XX

07-MAR-2002.

PD	07-MAR-2002.
XX	
PF	01-SEP-2001; 2001WO-EP10074.

XX	01-SEP-2000; 2000DE-1043826.
PR	
PR	05-SEP-2000; 2000DE-1044543.

PR	01-SEP-2000; 2000DE-1043826.
PR	05-SEP-2000; 2000DE-1044543.
XX	

PA	(EPIC-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Bern

PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.

PT	Determining the degree
PT	for diagnosis and pr
XX	
XX	

PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -

XX	comparisons from chemically treated DNA
XX	
PS	Claim 12; 56pp + Sequence Listing; 56pp; German
XX	

PS Claim 12; 56pp + Sequence Listing; 56pp; German

CC methylation of a particular cytosine in a motif

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.

CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the

CC member, if oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridization to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridized to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABO13410-ABQ3411 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

CC classified oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC AB013410-AB049141 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX
XX
XX Sequence 1447 BP; 654 A; 306 C; 134 G; 353 T; 0 other;

Sequence 1447 BP; 353 A; 134 C; 306 G; 654 T; 0 other;

Best Local Similarity 90.08; Pred. No. 72;

Query Match	Score	DB	Length
76.48;	16.8;	24;	1447;

```
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

[illegible]

3 TATACTACATTTCGAACGA 22

CC provided.
XX Sequence 25 BP; 12 A; 6 C; 0 G; 7 T; 0 other;
SQ

Query Match 76.4%; Score 16.8; DB 21; Length 25;
Best Local Similarity 90.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22
|||||
DB 6 TATATCTACATTCGAACGA 25

RESULT 6
ABQ42138/c
ID ABQ42138 standard; DNA; 522 BP.
XX
XX ABQ42138;
AC
XX
XX 12-JUL-2002 (first entry)
DT
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 28729.
DE
XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200218632-A2.
PN
XX
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP10074.
PF
XX
XX 01-SEP-2000; 2000DE-1043826.
PR
XX 05-SEP-2000; 2000DE-1044543.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI
XX
XX WPI; 2002-371829/40.
DR
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
PS
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX
XX Sequence 522 BP; 64 A; 70 C; 244 G; 141 T; 3 other;
SQ

Query Match 76.4%; Score 16.8; DB 24; Length 522;
Best Local Similarity 90.0%; Pred. No. 67;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TATATCTACATTCGAACGA 22
|||||
DB 27 TATATCTACATTCGAACGA 8

RESULT 7
ABQ42139
ID ABQ42139 standard; DNA; 522 BP.
XX
XX ABQ42139;
AC
XX
XX 12-JUL-2002 (first entry)
DT
XX
XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 28730.
DE
XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200218632-A2.
PN
XX
XX 07-MAR-2002.
PD
XX
XX 01-SEP-2001; 2001WO-EP10074.
PF
XX
XX 01-SEP-2000; 2000DE-1043826.
PR
XX 05-SEP-2000; 2000DE-1044543.
PR
XX
XX (EPIG-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI
XX
XX WPI; 2002-371829/40.
DR
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
PS
XX
XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX
XX Sequence 522 BP; 141 A; 244 C; 70 G; 64 T; 3 other;
SQ

Query Match 76.4%; Score 16.8; DB 24; Length 522;
Best Local Similarity 90.0%; Pred. No. 67;


```

/product="hypothetical protein"
/protein_id="AAD23041.1"
/db_xref="GI:4559381"

/translation="MAESSTIARLYRCCKKNLQEPKDCPPFOCGGCGTYLCAKNK
DRADLLSVKSVENRKELEVTSSEESDSVASLRHODGQTEWVDVHNDPSKMN
VDSGNDSQPGLDQIRKRTKRCDEGRFSTSNFTDLSLDEAIEQDRAGLVKRL
DKLEQVOPKEQOIPSSSSGQKAPLRYFSSGRYAGPSYQYSEPEPCLYNNNS
DVLHGPMHNPAVAYGDPHGFPMHETLQPSHLNSROYIGNNGHVDFTPHNGM
FHQSTCSWLPDALRNPGFYPHERSEGFSTLSHPRTFPPGSQSPGOLRGRTSGFK
DMEMNAVSRYHPKSPYSSSGSRLQPVAGAPETINCPEKLIKPEKIDSATRKO
RMRCGACSCYIDISFYDKLLISTDPASARKPETHRLRWETTANPSSDDYDNAIEE
HADRGSADVSTALVLISNTAQMQLTDAISPSVSEDELSDSSTVRAKYNLASPLAKQ
FEVSSINVRDSGPRQSSRSQEDRYTLSTKTAQRMQSMKEASVAIEMDVNDYSHNEV
SODSANDCTDDQGRTKKGGFASIMKNSFKDLKSIQNEGSDVSIINGHVAEDLYKMA
EKQAGPIRPGNMYDYRAGFWGLGSHCLGILPPFIEELNYPMPENCAGGTRVFNQ

```

```

Query Match      78.2%; Score 17.2; DB 8; Length 95214;
Best Local Similarity 86.4%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 CGTATATCTACATTGGAACGA 22
          |||||
Db      82489 CTTATATCTGCATTGGAACGA 82510

```

Search completed: July 8, 2003, 04:02:41
 Job time : 83.3011 secs

```

source      1. .95214
            /organism="Arabidopsis thaliana"
            /cultivar="Columbia"
            /db_xref="taxon:3702"
            /chromosome="2"
            /map="C1C02E07"
            /clone="F11C10"
            /complement(1. .7273)
            /note="overlap with BAC clone T3F17 (AC005397:1. .7273)."
```

gene

```

            /complement(2249. .2960)
            /gene="At2g46330"
            /note="F11C10.2; supported by cDNA:
            g1_15294169_gb_Af410276_1_Af410276"
            /complement(join(<2249. .2542,2700. .22960))
            /gene="At2g46330"
            2353. .2461
```

repeat_region

```

            /rpt_family="(GAA)n"
            /complement(join(2436. .2542,2700. .2814))
            /gene="At2g46330"
            /codon_start=1
            /product="expressed protein"
            /protein_id="AAD23036.1"
            /db_xref="GI:4559376"
            /translation="MASRNSVTGFRLEFSPFAVILSLAGASLAPAPAPISDGSIDQ
            GAIDLMVVALVTLIHLHDLKSSSTFF"
```

gene

```

            /rpt_family="Ar_rich"
            /complement(3425. .3473)
            /complement(6278. .8483)
            /gene="At2g46340"
            /note="F11C10.3"
            /complement(join(<6278. .6469,6630. .6689,6777. .7094,
            7253. .7399,7602. .>8483))
            /gene="At2g46340"
            /codon_start=1
            /product="putative photomorphogenesis repressor protein"
```

repeat_region

```

            /protein_id="AAD23037.1"
            /db_xref="GI:4559377"
            /translation="MACYCVNYLVSPFVLLIFLSCDRDILKSELICEDDSVKSTAA
            EISILLHLFSLSEVOKKRSKLLDIOCTLEDIOKEAERYSNVSLVSHGAIK
            ROSSPDIHECHTSALFPYPAATMDRLMSNTROLEDAFEPKRSQNTLSSAATASIK
            TLKDRDRCSENONENQDMSTKSSDQLEVFPEGLCKARISKETCTITSGDILNS
            ASVQSLSPDEEHLIAAGISKIKITIFENAFMESVGVHPLVEMVNRKSLSCVCM
            NSYIKNYLASTIDYGVQIWDAGTGOQSSOYTEHQRKMSVDFSPDTRKFSGDDC
            SVKLMSINEKRSLGITWSPANVCYQFESSYNHLIAFGSADYKVCYDLRKYKPMWCT
            LAGHKAVSYVKFMDSETIVASQTSLSKLNMLNKTINSSGLSPGACSLTYGHTMOKN
            FVGLSVLDGYIAGSETNEVYSYKSLPMPMTSYFGSVDPISGNEYRDDNGQFVSSV
            CMRRKSNMLVAANSSTGNMKILTV"
            /complement(9088. .10527)
            /gene="At2g46350"
            /note="F11C10.4; predicted by genefinder"
            /complement(join(<9088. .9392,9441. .>10527))
            /gene="At2g46350"
            /complement(join(9088. .9392,9441. .10527))
            /gene="At2g46350"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAD23038.1"
            /db_xref="GI:4559378"
            /translation="MPYMERVAEEVATVNTNIOAKARVDVPCNKLDARHNDVIOSET
            AASDPSGAHRNVOLTKRPPEBAAGAKLSYEELTIGNVRYOGSNNTNPPSPAGK
            FHLIRLARGSAFRAGDGDIDSDPRDMDLSRIQOULGAPSEKONTLKPFRSRSDO
            NLEAFSERLAAGENSINAPALISEGVQMKRPVSSNSFSLLLKRAKRGKRVAKNO
            EEPPEFVSDDQSKSEKLDISKSPHIVLPLKSPKNGVSHGDNHSSGIS
            LREPLFSSAKREKRHGLCLFQLVLELVAISKSLFLDLRPSLFTLPSPSKLRYIG
            NGKNDLESDVEDLNRPRPVYESSSGNQOATSTGPRFKRSPVIDLNNVDARNP
            DSCPLQODDYITKNLSVSSYSRKSQSMSTWLEEWYTCPEPINEGDIIGESNITAYALCVLL
            EEWNTQPYCCG"
```

repeat_region

```

            /protein_id="10661"
            /rpt_family="(GAA)n"
            10663. .10661
            repeat_region
            10966. .10990
```

gene

```

            /rpt_family="(GAA)n"
            12213. .12506
            /gene="At2g46360"
            /note="F11C10.5; predicted by genefinder"
            <12213. .>12506
            /gene="At2g46360"
            12213. .12506
            /gene="At2g46360"
            /codon_start=1
            /product="hypothetical protein"
            /protein_id="AAD23039.1"
            /db_xref="GI:4559379"
            /translation="MDSREANRKHDAEERREDEKTSLSAEVNIAMVLLSOLSTKKEDE
            NGLEKERDVAHAEARSRARANKREFKDKTLMEKRTHLVQPAKEFRHYRKM"
```

CDS

repeat_region

```

            /rpt_family="AT_rich"
            14702. .14733
            /rpt_family="(TA)n"
            17495. .20199
            /gene="At2g46370"
            /note="F11C10.6; supported by cDNA:
            g1_15292854_gb_Af050861_1"
            join(17495. .17656,18047. .18346,18414. .18515,18662. .19449,
            19537. .20199)
            /gene="At2g46370"
            /complement(17508. .17548)
            /rpt_family="(GA)n"
            /complement(17871. .17892)
            /rpt_family="AT_rich"
            join(18048. .18346,18414. .18515,18662. .19449,19537. .20075)
            /gene="At2g46370"
            /codon_start=1
            /product="putative auxin-responsive protein"
            /protein_id="AAD23040.2"
            /db_xref="GI:20197821"
            /translation="MLEKVEFPDKNRVYIDFEDENTRAHVOXQTLKEILLKQNSAIV
            LONCGINGNADPEERAKSVPLVTDVLEPPYIKRMVDGDTSPILGPPVPAISLGG
            TSGCRPKRIFPTDLEMTNIOLEPRTAFARNRDEPIDNRKALQFITSSQYISTGV
            PVGATATTVNVRNPNFKAGKMSITSPSCPEVITSPVHQAALCHLSGLFLEDOVY
            VFAFAGLVAHAFRTPEQVWEIVTDIKDGLSNRIYVPSVKRAMSKLLPNEPLEAT
            IRTCAFLSNMVGILPAPLPNARYGIMKDSMEPPYKLRHVGADLPLVSHHSGSE
            GMIANVTPLRSPPEATFAVIPNLYGEFFLPVSTGGEKRPVLTQVRIKGEYEVAT
            TNVAGLRVYRGLDPAKYGIVGYNMPLQKFCRRLLISTINDKTERDQLSESAK
            RLSEKKEFVIDESYIDVSTDPRGATAFMEIISGFTNEDVLODCNCIDRAREFTAGVVS
            SRKKTIGALELRVANGTFRKIOEHPLGLGSSRQGFMRCPVPSNAKVLQILCENV
            VSSYFSRTAF"
            21483. .21515
            /rpt_family="(CA)n"
            21570. .22086
            /gene="At2g46375"
            /note="F11C10.29; supported by full length cDNA: Ceres:
            94104"
            21570. .22086
            /gene="At2g46375"
            21664. .22035
            /gene="At2g46375"
            /codon_start=1
            /product="expressed protein"
            /protein_id="AAM15262.1"
            /db_xref="GI:20197819"
            /translation="MPNTRILGSSSEERKISDGRDPKIMDCSTLYDSYELVSEFVHIE
            RKLPFSLAQPGITLRALMDKXSDHCASCSTRGPCIHRRKYWMGKKKDEKERTI
            NKKAMPDSFWMNCKNLF"
```

gene

```

            23083. .25732
            /gene="At2g46380"
            /note="F11C10.7; predicted by genefinder"
            join(<23083. .23203,23417. .24145,24203. .25262,25346. .25406,
            25505. .>25732)
            /gene="At2g46380"
            join(23083. .23203,23417. .24145,24203. .25262,25346. .25406,
            25505. .25732)
            /gene="At2g46380"
            /codon_start=1
```

CDS

```

TAFEIGTTCGCTTGGKSLVEEVEEAQEAVALPGTSEAFLEALSVDVNDYHLHDSOEH
"
complement(14394..15192)
/gene="At2g46505"
/notes="F13A10.20; supported by full length cDNA: Ceres:
11006"
complement(<14394..>15192)
/gene="At2g46505"
complement(14656..15111)
/gene="At2g46505"
/codon_start=1
/product="Expressed protein"
/protein_id="AA15243.1"
/db_xref="GI:20197780"
/translation="MSLRRTLDLHROTGRATLSKSPESMSHSSASATAVRNPLG
RLSSITPFAOAKIKKSDSNVYTRDSISSGSSKLNAAKRSRSSRGYTNSEFLR
KIPYVPHIESMEELIADYVHEGEMRNLYMSLGLFQIIVLKDILFLT"
complement(15830..17530)
/gene="At2g46510"
/notes="F13A10.4"
complement(<15830..>17530)
/gene="At2g46510"
complement(15830..17530)
/gene="At2g46510"
/codon_start=1
/product="putative bHLH transcription factor"
/protein_id="AA20162.2"
/db_xref="GI:20197775"
/translation="NMNMDLGDWDEDKSVSAVGLHSLASPLRANSNONTFLWGT
DTLNKLSLVDMPSSEFNRYALFROQITRSRGQVYLGDCGCEPNEEESKV
VRSYNNNGAESEITWQDKRKYOLKHLRFGSDNDALSLEKVTATEIFLASM
FEFNHGGGGRGYSCKHVLSDAVNSDCCFRFSKASGIRTIWVPDAGVLE
LGRVMSLPIENIGLVKSVQALFMRVYIQPMVTSNMGIGIKHFGODISGAHAPK
LEVRNLDERETPOSMEGYNKNGKPTFGTPORDVYVENVNMYDNNKXTOLEA
GSSVAASNPSTNTOEKSCEKRPVSLAGAGIVSVDEKRPKRPANRE
PLNHVABORREKLNORVYALRSVYPTSKMDKASITGATSYRKEIOEKXIMDE
RNGCTKSIESNTITVESEPEVDIQAMEEYVRYISPLDSHPARITIQAMNSVSL
MEAKLSLADIMEHTFVIVKSNMGSDPLTEKLIAPYPTSTQPLPSSSSQVSGDI
"
repeat_region
17920..17946
/rpt_family="(GA)n"
complement(18392..18454)
/rpt_family="AT_rich"
21510..24428
/gene="At2g46520"
/notes="F13A10.5"
<21510..>24428
/gene="At2g46520"
21510..24428
/gene="At2g46520"
/codon_start=1
/product="putative cellular apoptosis susceptibility
protein"
/protein_id="AA20163.1"
/db_xref="GI:4415933"
/translation="MEMRETLVETLSQCLNTLSPIPEPRTAERALSAAOANGL
AVRLVAEPALDQCTRHAAVNRKNLBRMRHAGSGISPYDSKEQIKITVISM
LSASPIQSOLESALVYIGKHPKAMPALPELNLALAGVSVNLTIGLAS
STKRTSYERIDALFDLKYCLDNLAAPLTEFLTSSLLDASGSGSPFLPLP
ESORLCTITFVNFQDLPEFEDNKKEMGKRTKLSNYPALSETEGTLVLDPL
AAICENINHYIENEEFOGLNEFASVWTLIDVSKSPRSROSLTKLSTRT
SVHHLFAGDNVYKEICOSIVPNVSLRVEDIEIFMNYIEFRRMESDQVTRST
ACELLGATNRYKTYEYVSLIORLISFSANPKMKDKCAIYIVYSTKAG
GASVSTDLIDVQNFANITLPELOSDVNSPMLKAGSLKFLTFMESHLPKPAQOL
PELVRLKESNVSHYASCIKELLVVEEGRGNRYAGDLSPLDLMNLDAL
KPESEENYIANKIMRVGVADISAENVAGPCIGLITLSEVCKNKPINNHLE
SVAVLVRACERDILISAFETSLPQLMIANDITEPLPYGQLADVLNPLT
SPVYNOIFILLSPESMRSGNVALRLLQAFLOKAPHEVQENKLSVIEKLV
ASPTDEGFEYIINTIENDSVIAVYKGVASALFTRQNKTKVKKPOKSVIFKSL
FLVKGCAVIVETMTVQPNITIAIVHEWIRPKIKMSMEYKFLAVATILICETP
ALIDPSAKIKMGKMLDSIVTLVSRPEORVLDPEPELISENGITAAVAKHNGKK
EEDPLKIDKOPKQFLVASVRSLSASPGRRPQIIGENLEAOTATLQICNAYNGCIA
"

```

```

repeat_region 25088..25183
/rpt_family="Rf:g11100803|emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"
repeat_region 25090..25183
/rpt_family="(GA)n"
25090..25166
/rpt_family="Rf:g11100803|emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"
repeat_region 25090..25162
/rpt_family="Rf:g11100803|emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"
repeat_region 25090..25158
/rpt_family="Rf:g11100803|emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"
repeat_region 25090..25154
/rpt_family="Rf:g11100803|emb|X93607.1|AT74A.A.thaliana
microsatellite repeat region (clone 74A)#Rf:ATR0046|X93607
74A"

```

```

Query Match 78.2%; Score 17.2; DB 8; Length 84825;
Best Local Similarity 86.4%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 CGTATCTCATTCGAAACGA 22
Db 23427 CTTATATCTGCATTGAAACGA 23448

```

```

RESULT 15
AC006526
LOCUS
DEFINITION
Arabidopsis thaliana chromosome 2 clone F1C10 map C1C02E07,
complete sequence.
ACCESSION
AC006526
VERSION
AC006526.8 GI:20197817
KEYWORDS
HTG.
SOURCE
Arabidopsis thaliana.
ORGANISM
Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 95214)
Lin.X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
Barnstead,M.E., Mason,T.M., Bowman,C.L., Renning,C.M.,
Benito,M., I., Carrera,A.J., Creasy,T.H., Bell,C.R., Town,C.D.,
Nierman,W.C., Fraser,C.M. and Venter,J.C.
Unpublished
2 (bases 1 to 95214)
Lin.X.

```

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA, cdtown@igf.org
On Apr 18, 2002 this sequence version replaced gi:6598573.
On Oct 6, 1997 this sequence version replaced gi:2443866.
We have determined that YAC YUP812 is chimeric, and is comprised
of two distinct genomic EcoRI fragments from chromosome 1. This
submission contains the sequence from the EcoRI site at position 1
(right end) to position 181918 of our previous Phase II
submission YUP812 accession number AC000098. This fragment maps
at the bottom of the right arm of the chromosome between the ATHTAPS
and m532 markers.
Location/Qualifiers

```

FEATURES

Query Match 78.2%; Score 17.2; DB 8; Length 82454;
 Best Local Similarity 86.4%; Pred. No. 5.4e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

gene 1 CGTATACCTACATCGAAGCA 22
 ||||| ||||| ||||| ||||| |||||
 Db 36634 CGTATATATACGTTCAAGCA 36655

RESULT 14
 LOCUS AC006418 84825 bp DNA linear PLN 11-MAR-2002
 DEFINITION Arabidopsis thaliana chromosome 2 clone F13a10 map C1C06C03,
 complete sequence.
 ACCESSION AC006418
 VERSION AC006418.4 GI:20197771
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 84825)
 Lin, X., Kaul, S., Shee, T. P., Fujii, C. Y., Shen, M., Vanaken, S. E.,
 Barnstead, M. E., Mason, T. M., Bowman, C. L., Rongman, C. M.,
 Beitel, M. J., Carrera, A. J., Cressy, T. H., Buell, C. R., Town, C. D.,
 Nierman, W. C., Fraser, C. M. and Ventner, J. C.
 Unpublished
 2 (bases 1 to 84825)
 Lin, X.
 Direct Submission
 Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 3 (bases 1 to 84825)
 Town, C. D. and Kaul, S.
 Direct Submission
 Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
 On Apr 18, 2002 this sequence version replaced gi:5698561.
 Location/Qualifiers
 1. 84825
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"

misc_feature
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C06C03"
 /clone="F13a10"
 1..36153
 /note="overlap with BAC clone F11C10
 (AC006526:59064..95214)."

gene
 /note="overlap with BAC clone F11C10
 (AC006526:59064..95214)."
 complement(1047..3029)
 /gene="At2g46450"
 /note="F13a10.1; predicted by genescan"
 complement(join(<1047..1322,1409..1518,1595..1997,
 2074..2431,2503..2828,2916..3029))
 /gene="At2g46480"
 complement(join(1047..1322,1409..1518,1595..1997,
 2074..2431,2503..2828,2916..3029))
 /gene="At2g46480"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA020159.1"
 /db_xref="GI:4415931"
 /translation="MTDCCCKGNDKVPVPRGCTGTGKAFNDPTPELHRSIRLRQK
 RIERANELNDSLCKLFTLAARSRVSDAPLGVITWKEVIRGSSFEDMLWD
 QIIMARVYGLGKLFNNALHOEITQVMKLAWESESTDIDQEVRLDSIDMKQIILA
 RAHEOLYBECKLVNKLRLAMLOVEDELENOETITFLTLQSKALPDAICLTPLRLV
 EYHLPLPRNPRLNENPLKLYALFSDNLAASSVSTVMNODPSRHYEHLV
 TSKINGGSMFELNPGEARITHVOREPDSNVLAASVPSLQLESAMKKEFKTA
 RSESVSGSENLKIRPKYMSMLNLRIIRIPFKELILFVDDVYOKDLTPWS
 IDIKGVNENFDKFGVAGVGNITFDLEMKNNITETIHYMWNENRILAKIGLTP
 PGLITFVNLTPQLQRKMLLGLGDKDGIVAKIERSAVIHVGHAKPTEWGISKYP
 WYKYTFNDHPYIFICRLFE"
 4060..4696
 /gene="At2g46490"
 /note="F13a10.2"
 4060..4696
 /gene="At2g46490"
 4100..4504
 /gene="At2g46490"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AA020160.1"
 /db_xref="GI:4415930"
 /translation="MAADIFRSIFEGCISGDSIAEERPYHKNGCALHDKSGGACK
 NQNRPPSPCRRHSGSEISFPIRRSSGCKTAMNPLSSSSSSNLOSLSSSSSLSN
 LASDLPVNAAETPEPSRSKQRLWTIDEEDD"
 complement(4483..4517)
 /rpt_family="CAZn"
 complement(6756..6792)
 /rpt_family="Al_rich"
 complement(11207..13602)
 /gene="At2g46500"
 /note="F13a10.3; similarity to ubiquitin family of
 proteins; supported by cDNA: gi_16930424_gb_Af419566
 1..Af419566"
 /gene="At2g46500"
 complement(join(<11207..11764,11858..13602))
 /gene="At2g46500"
 /codon_start=1
 /product="expressed protein"
 /protein_id="AA020161.1"
 /db_xref="GI:4415931"
 /translation="MSSAGVALSPVRSFELMPLVANSCLDSYDDDTIMYLTLPFG
 VIMRVLESDSIESVKLRIOYSRGVARNQKLVFGRELARNSNMRYGVSEGNILH
 LVKLSDIQLVDVTKTQGHCRHFVRGNIGVYKQISKRGGEDVDEDEILYEGG
 KLEDQSLINICNDSDSVLILVRSAKVRVPRKNEFLSIAPQADKGRKSKIS
 VPKRISLEPVVNSAKVPLVYKMDIQSAGLSGNSPVSSGCGAVEMQPSGSG
 NKRVGKFPIDEPMAENNPGLPLSPNGEGIKKGTXYGGLRPAVAYITIDHPSGG
 KSMFGEIGFAGVPTAMIECLHPGNHPKGIKRTIGISLQNTENDSCDEMGPSF
 VEEVHKISVLDILANADRHGGLNLTWKESKSLVLPIDHGYCLPEFEDCTEWL
 WPARVKSATDEYIRSDAEDIDILKFGMKKPAETLAQTLRISTYLKKGVEGRIG

misc_feature
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="C1C06C03"
 /clone="F13a10"
 1..36153
 /note="overlap with BAC clone F11C10
 (AC006526:59064..95214)."

gene
 /note="overlap with BAC clone F11C10
 (AC006526:59064..95214)."
 complement(1047..3029)
 /gene="At2g46450"
 /note="F13a10.1; predicted by genescan"
 complement(join(<1047..1322,1409..1518,1595..1997,
 2074..2431,2503..2828,2916..3029))
 /gene="At2g46480"
 complement(join(1047..1322,1409..1518,1595..1997,
 2074..2431,2503..2828,2916..3029))
 /gene="At2g46480"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA020159.1"
 /db_xref="GI:4415931"
 /translation="MTDCCCKGNDKVPVPRGCTGTGKAFNDPTPELHRSIRLRQK
 RIERANELNDSLCKLFTLAARSRVSDAPLGVITWKEVIRGSSFEDMLWD
 QIIMARVYGLGKLFNNALHOEITQVMKLAWESESTDIDQEVRLDSIDMKQIILA
 RAHEOLYBECKLVNKLRLAMLOVEDELENOETITFLTLQSKALPDAICLTPLRLV
 EYHLPLPRNPRLNENPLKLYALFSDNLAASSVSTVMNODPSRHYEHLV
 TSKINGGSMFELNPGEARITHVOREPDSNVLAASVPSLQLESAMKKEFKTA
 RSESVSGSENLKIRPKYMSMLNLRIIRIPFKELILFVDDVYOKDLTPWS
 IDIKGVNENFDKFGVAGVGNITFDLEMKNNITETIHYMWNENRILAKIGLTP
 PGLITFVNLTPQLQRKMLLGLGDKDGIVAKIERSAVIHVGHAKPTEWGISKYP
 WYKYTFNDHPYIFICRLFE"
 4060..4696
 /gene="At2g46490"
 /note="F13a10.2"
 4060..4696
 /gene="At2g46490"
 4100..4504
 /gene="At2g46490"
 /codon_start=1
 /product="unknown protein"
 /protein_id="AA020160.1"
 /db_xref="GI:4415930"
 /translation="MAADIFRSIFEGCISGDSIAEERPYHKNGCALHDKSGGACK
 NQNRPPSPCRRHSGSEISFPIRRSSGCKTAMNPLSSSSSSNLOSLSSSSSLSN
 LASDLPVNAAETPEPSRSKQRLWTIDEEDD"
 complement(4483..4517)
 /rpt_family="CAZn"
 complement(6756..6792)
 /rpt_family="Al_rich"
 complement(11207..13602)
 /gene="At2g46500"
 /note="F13a10.3; similarity to ubiquitin family of
 proteins; supported by cDNA: gi_16930424_gb_Af419566
 1..Af419566"
 /gene="At2g46500"
 complement(join(<11207..11764,11858..13602))
 /gene="At2g46500"
 /codon_start=1
 /product="expressed protein"
 /protein_id="AA020161.1"
 /db_xref="GI:4415931"
 /translation="MSSAGVALSPVRSFELMPLVANSCLDSYDDDTIMYLTLPFG
 VIMRVLESDSIESVKLRIOYSRGVARNQKLVFGRELARNSNMRYGVSEGNILH
 LVKLSDIQLVDVTKTQGHCRHFVRGNIGVYKQISKRGGEDVDEDEILYEGG
 KLEDQSLINICNDSDSVLILVRSAKVRVPRKNEFLSIAPQADKGRKSKIS
 VPKRISLEPVVNSAKVPLVYKMDIQSAGLSGNSPVSSGCGAVEMQPSGSG
 NKRVGKFPIDEPMAENNPGLPLSPNGEGIKKGTXYGGLRPAVAYITIDHPSGG
 KSMFGEIGFAGVPTAMIECLHPGNHPKGIKRTIGISLQNTENDSCDEMGPSF
 VEEVHKISVLDILANADRHGGLNLTWKESKSLVLPIDHGYCLPEFEDCTEWL
 WPARVKSATDEYIRSDAEDIDILKFGMKKPAETLAQTLRISTYLKKGVEGRIG

/product="ribosomal protein s18 (rpsr)"
 /protein_id="AAC65056.1"
 /db_xref="GI:3322319"
 /translation="MAEDHPSVDLTHISSPSESESAKKNQFRRKKVCEPCQKL
 LADYKPDILRRITERRGILPRITIGTCARHORVALEKRSRAVALLPVLITE"
 complement(2593) .3123)
 /gene="TP0062"
 complement(2593) .3123)
 /note="similar to GB:AE000783 percent identity: 48.15;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="single-strand DNA binding protein (ssb)"
 /protein_id="AAC65057.1"
 /db_xref="GI:3322320"
 /translation="MADVNRVYVGLRTDAELKITSAGALCFESVAINRRKSGD
 WVEVNEFDIVLMRGQEVTSYLTKGQVAVGEGLQKRWEDGSRKVELSRINV
 QILGSVYGARADEGEFSSRYAAESTAGRGVTSDDSRPTDILGKKRGIDATSS
 LDADFSSTDLDTVP"
 complement(3132) .3413)
 /gene="TP0063"
 complement(3132) .3413)
 /note="similar to GB:AE000783 percent identity: 30.00;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="ribosomal protein s6 (rpsf)"
 /protein_id="AAC65058.1"
 /db_xref="GI:3322321"
 /translation="MRTYELMAVSAHEDLFIQSGAVRALLQENDAVIAREDHIGER
 ELAVPLKKRGKRYLLFTVQCEPKVRELDHKLRLHDLTHLFVYVDS"
 3555 .4130
 /gene="TP0064"
 3555 .4130
 /gene="TP0064"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0064"
 /protein_id="AAC65064.1"
 /db_xref="GI:3322327"
 /translation="MLPTAQRRLRGYTRACAPYGRGEEPPVRRCAFLCARVGHVCG
 RSVYQNFQDNIYHLARSIDVYEGQLMDCLYAEKVVYDVEFDHALQCLAH
 DRCHPEVDYPLICLFCGCAFLNLNLFHRAAPVTAQVEFTMLPFEKRRSHALH
 ENLARSIOEVDTSADASHVVSODEIDELLEH"
 4204 .4791
 /gene="TP0065"
 4204 .4791
 /gene="TP0065"
 /note="similar to GP:1694865 percent identity: 36.11;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65059.1"
 /db_xref="GI:3322322"
 /translation="MLKNHYLRCPDGPTRPAMDRESLEFALLGDMGSCFLDLFAS
 GVGLEAVSRGAYVVFVEMVRSFVLLQVVALGLERCQAVARYATARATLTH
 FVYLDPPEPVRFHAEILLQRTSRASICREGSVVMVHRRREKKLADKIDSLVRTQRYVGS
 RSVYDFRIRIKPVRRTEIRCRILLFVRGDAEGA"
 4775 .5089
 /gene="TP0066"
 4775 .5089
 /gene="TP0066"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0066"
 /protein_id="AAC65065.1"

/db_xref="GI:3322328"
 /translation="MORVLESDFYFVNGIGRPVSTLSDRDLALLRNGAYLNEGLK
 CEARVFTITGHDGLTRIGDYMKADVLTALRFYFAENQKMRPIVASLVLIRIC
 LI"
 5077 .6258
 /gene="TP0067"
 5077 .6258
 /note="similar to GB:AE000783 percent identity: 47.91;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65060.1"
 /db_xref="GI:3322323"
 /translation="MSYIILTPGGGMLGCECSVPPEFVETGKRSVDADRIMLSR
 DGARLKYNDELRASEGKILIOIADNNVALVAGDAARARRAYOEASDYTRCLQ
 YRNSYALFGLDLCYKNNRRYRKAYEVMOCYEDDSHNIATVTPADAYRKIDHDS
 NUYISOVLNDEHNSYALIGLAHLHPDCKRREALITYKKLLCEASHVDIVLITSG
 NCYRRMKLEFQGLPYFQELKRPDGNFYGFQFMADCYGMKNQENSIYWERILEKDI
 ONRVILITRADAYRAHIGEXKAHQYKRALDIDYAVATGLAVLCIKLQGREAVVS
 LDRLVQLDKKNRYRVEVLADCYRKLQKQKALETLRPPQDFGKNNRVSELSNELEGA
 S"
 5439 .7461
 /gene="TP0068"
 5439 .7461
 /gene="TP0068"
 /note="similar to PID:1001734 PID:1208474 percent
 identity: 39.39; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65061.1"
 /db_xref="GI:3322324"
 /translation="MEMCCALSGLLPEIOLKVCAPAFERRRGVQVFRMIAAGCTDEHAM

Query Match 78.2%; Score 17.2; DB 1; Length 11101;
 Best Local Similarity 86.4%; Pred. No. 5.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	1	CGATATCTACATTCGAACGA	22
Db	1		
	4467	CATACATCTACATTCGAACGA	4446

RESULT 13
 AC012193
 LOCUS
 DEFINITION
 complete sequence.
 AC012193
 AC012193.6
 GI:12323286
 HTG.
 Arabidopsis thaliana.
 Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 82454)
 Lin.X., Kaul, S., Town, C.D., Beattie, M., Creasy, T.H., Haas, B.J.,
 Wu, D., Maitl, R., Rounsley, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
 Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
 Arabidopsis thaliana chromosome 1 BAC T3228 genomic sequence
 Unpublished
 2 (bases 1 to 82454)
 Lin.X. and Kaul, S.
 Direct Submission
 Submitted (21-OCT-1999) The Institute for Genomic Research, 9712
 Medical Center Dr. Rockville, MD 20850, USA, xlnet@ig.ing
 3 (bases 1 to 82454)
 Town, C.D. and Kaul, S.
 Direct Submission

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE

REFERENCE
 MEDLINE
 98332770
 9665876
 2 (bases 1 to 11101)
 Fraser,C.M., Norris,S.J., Weinstock,G.M., White,O., Sutton,G.G.,
 Dodson,R., Gwin,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
 Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
 Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
 Uterback,T., McDonald,L., Artlich,P., Bowman,C., Cotton,M.D.,
 Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Matthey,L.,
 Weidman,J., Smith,H.O. and Venter,J.C.
 Direct Submission
 Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 location/Qualifiers
 1..11101
 /organism="Treponema pallidum"
 /db_xref="taxon:160"
 complement(68..1384)
 /gene="TP0058"
 complement(68..1384)
 /gene="TP0058"
 complement(68..1384)
 /note="similar to GB:D26185.SP:P37469.PID:467330
 GB:AL009126 percent identity: 39.68; identified by
 sequence similarity; putative"
 /codon_start=1
 /translation="replicative DNA helicase (dnab)"
 /protein_id="AAC65054.1"
 /db_xref="GI:3322317"
 /translation="MFGKPNFTQELKGIPEPHNEAEKAVGLNDDLSALSTEDL
 SASSFTSAHQRIQALVELSDGQRPDILVLSHLSCEALDFVGSATVASTIDV
 PSAANVEYTRIVCDAMRSLKVARITAEANDVSGNIVETQREYVETDAR
 RVATFKLKNLIPDLVNIETRYRNSQSLVATGLDALDNLGTGPNSEIYVIGAP
 MNGFKALMTASNAIQRIPRTAFESLNSNLALMORLNAEGVGSATNLKGLDL
 SDFGIQAAGEWDAPLTYVDVNMKLDIAVARRICVGERIQITFVYIGIVAD
 NPAPRYQPAISQISKLARELDIPVALSGVGRAPGSAAPLADIRSGALEQDA
 DVNMFHARDNETQTLAKQKRNPIGTVELEFQASFTREVCNSP"
 complement(1368..1595)
 /gene="TP0059"
 complement(1368..1595)
 /gene="TP0059"
 complement(1368..1595)
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /translation="T. pallidum predicted coding region TP0059"
 /protein_id="AAC65063.1"
 /db_xref="GI:3322326"
 /translation="MRKYLARSMSGCSFSCAKNKATLIGDSTEGRFVAVGPFFAS
 LEFRHSGDCLFTTFPLSTPRRRCFAGHA"
 complement(1601..2071)
 /gene="TP0060"
 complement(1601..2071)
 /gene="TP0060"
 complement(1601..2071)
 /note="similar to GB:D26185.SP:P37437.PID:467335
 GB:AL009126 percent identity: 32.67; identified by
 sequence similarity; putative"
 /codon_start=1
 /translation="ribosomal protein L9 (rpL1)"
 /protein_id="AAC65055.1"
 /db_xref="GI:3322318"
 /translation="MKITINDVKTIGSGGVKVAAGFENYLYPRLVAPNARFTV
 ARFKRQDDIERKSLKRODANLKAERLADPVIYAMAGNGLKAVTSHTVABOL
 ACMGFEVERKREYVGLTKCVGNHVTIRLYEDICAVPVIKNQSEADSVSE"
 complement(2177..2476)
 /gene="TP0061"
 complement(2177..2476)
 /note="similar to SP:P10806 percent identity: 51.95;
 identified by sequence similarity; putative"
 /codon_start=1
 /translation="

TITLE
 JOURNAL
 MEDLINE
 98332770
 9665876
 2 (bases 1 to 11101)
 Fraser,C.M., Norris,S.J., Weinstock,G.M., White,O., Sutton,G.G.,
 Dodson,R., Gwin,M., Hickey,E.K., Clayton,R., Ketchum,K.A.,
 Sodergren,E., Hardham,J.M., McLeod,M.P., Salzberg,S., Peterson,J.,
 Khalak,H., Richardson,D., Howell,J.K., Chidambaram,M.,
 Uterback,T., McDonald,L., Artlich,P., Bowman,C., Cotton,M.D.,
 Fujii,C., Garland,S., Hatch,B., Horst,K., Roberts,K., Matthey,L.,
 Weidman,J., Smith,H.O. and Venter,J.C.
 Direct Submission
 Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 location/Qualifiers
 1..11101
 /organism="Treponema pallidum"
 /db_xref="taxon:160"
 complement(68..1384)
 /gene="TP0058"
 complement(68..1384)
 /gene="TP0058"
 complement(68..1384)
 /note="similar to GB:D26185.SP:P37469.PID:467330
 GB:AL009126 percent identity: 39.68; identified by
 sequence similarity; putative"
 /codon_start=1
 /translation="replicative DNA helicase (dnab)"
 /protein_id="AAC65054.1"
 /db_xref="GI:3322317"
 /translation="MFGKPNFTQELKGIPEPHNEAEKAVGLNDDLSALSTEDL
 SASSFTSAHQRIQALVELSDGQRPDILVLSHLSCEALDFVGSATVASTIDV
 PSAANVEYTRIVCDAMRSLKVARITAEANDVSGNIVETQREYVETDAR
 RVATFKLKNLIPDLVNIETRYRNSQSLVATGLDALDNLGTGPNSEIYVIGAP
 MNGFKALMTASNAIQRIPRTAFESLNSNLALMORLNAEGVGSATNLKGLDL
 SDFGIQAAGEWDAPLTYVDVNMKLDIAVARRICVGERIQITFVYIGIVAD
 NPAPRYQPAISQISKLARELDIPVALSGVGRAPGSAAPLADIRSGALEQDA
 DVNMFHARDNETQTLAKQKRNPIGTVELEFQASFTREVCNSP"
 complement(1368..1595)
 /gene="TP0059"
 complement(1368..1595)
 /gene="TP0059"
 complement(1368..1595)
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /translation="T. pallidum predicted coding region TP0059"
 /protein_id="AAC65063.1"
 /db_xref="GI:3322326"
 /translation="MRKYLARSMSGCSFSCAKNKATLIGDSTEGRFVAVGPFFAS
 LEFRHSGDCLFTTFPLSTPRRRCFAGHA"
 complement(1601..2071)
 /gene="TP0060"
 complement(1601..2071)
 /gene="TP0060"
 complement(1601..2071)
 /note="similar to GB:D26185.SP:P37437.PID:467335
 GB:AL009126 percent identity: 32.67; identified by
 sequence similarity; putative"
 /codon_start=1
 /translation="ribosomal protein L9 (rpL1)"
 /protein_id="AAC65055.1"
 /db_xref="GI:3322318"
 /translation="MKITINDVKTIGSGGVKVAAGFENYLYPRLVAPNARFTV
 ARFKRQDDIERKSLKRODANLKAERLADPVIYAMAGNGLKAVTSHTVABOL
 ACMGFEVERKREYVGLTKCVGNHVTIRLYEDICAVPVIKNQSEADSVSE"
 complement(2177..2476)
 /gene="TP0061"
 complement(2177..2476)
 /note="similar to SP:P10806 percent identity: 51.95;
 identified by sequence similarity; putative"
 /codon_start=1
 /translation="

Query Match 80.9%; Score 17.8; DB 3; Length 300994;
 Best Local Similarity 90.5%; Pred. No. 2.6e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATCTACATCGAAGCA 22
 |||||
 Db 225380 GATACACTACATCGAATCGA 225360

RESULT 9
 AY093111 2438 bp mRNA linear PLN 21-APR-2002
 LOCUS Arabidopsis thaliana putative protein (At3g56140) mRNA, complete

DEFINITION cds.
 ACCESSION AY093111
 VERSION AY093111.1 GI:20260423
 KEYWORDS FFL.CDN.
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 2438)

Southwick, A., Katlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M.,
 Palm, C. J., Bowser, L., Jones, T., Banu, J., Carninci, P., Chen, H.,
 Cheuk, R., Chung, M. K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
 Kawai, J., Kim, C., Lin, J., Liu, S. X., Narusaka, M., Pham, P. K.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,
 Shinzaki, K., Ecker, J., Theologis, A. and Davis, R. W.

Direct Submission
 Submitted (27-MAR-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA

COMMENT e-mail for correspondence: arab@sequence.stanford.edu

RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : RIKEN
 Arabidopsis Full-Length cDNA) : Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinzaki, K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs. Southwick, A.,
 Nguyen, M., Katlin-Neumann, G., Lam, B., Miranda, M., Palm, C. J.,
 Bowser, L., Jones, T., Banu, J., Chen, H., Cheuk, R., Chung, M. K.,
 Kim, C., Lin, J., Liu, S. X., Pham, P. K., Sakano, H., Shinn, P.,
 Yamada, K., Ecker, J., Theologis, A. and Davis, R. W.

Southwick, A., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed
 equally to this work. Shinzaki, K. (RIKEN GSC) and Davis, R. W.
 (SSP/Stanford) contributed equally to this work as PIs.

FEATURES

source

gene

CDS

1..2438
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="3"
 /clone="RAFL09-27-C13 (R17974)"
 /note="This clone is in pBluescript
 ecotype: Columbia"
 1..2438
 /gene="At3g56140"
 /note="At3g56140"
 62..2299
 /gene="At3g56140"
 /codon_start=1
 /product="putative protein"
 /protein_id="AA013110.1"
 /db_xref="GI:20260424"
 /translation="MTITNDI.FA.SOP.LFHCPCSSSRPRRRRGIFRPIITLSEKN
 NSLSTALSDSDLPKRTASRAFLAPPLVASLFLKPSVSLASESSSRTITSP
 AVSAPPPPATITSPPPPKVEETITRTIDATIGAPPMAGDKDKRWKLLNAR
 VYLGEADVPPTKDKLELIVNRKRCVSEFOI SVALEAPFLDLODDQINQYMK
 RMDGETLKSYVTHWPAORWQVEEPLSTYCRDNSVRLIACGPTLKLRLTVOAGIGLS"

BASE COUNT 648 a 546 c 556 g 688 t
 ORIGIN

Query Match 78.2%; Score 17.2; DB 8; Length 2438;
 Best Local Similarity 86.4%; Pred. No. 6e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CGTATCTACATCGAAGCA 22
 |||||
 Db 959 CGTAGATCTACATCGAATCGA 980

RESULT 10
 ATH297282 2919 bp mRNA linear PLN 06-JAN-2002
 LOCUS Arabidopsis thaliana partial mRNA for cellular apoptosis

DEFINITION susceptibility protein homologue (CAS gene).
 ACCESSION AT297282
 VERSION AT297282.1 GI:18077709

KEYWORDS cas gene; cellular apoptosis susceptibility protein homologue.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

Location/Qualifiers
 1..2919
 /organism="Arabidopsis thaliana"
 /variety="ecotype Col-0"
 /db_xref="taxon:3702"
 1..2916
 /gene="CAS"
 <1..2916
 /gene="CAS"
 /function="nuclear export receptor for importin alpha"
 /note="CAS interacts with importin alpha 1-4 and with ran
 from Arabidopsis thaliana"
 /codon_start=1
 /evidence="experimental"
 /product="cellular apoptosis susceptibility protein
 homologue"
 /protein_id="CAC83300.1"
 /db_xref="GI:18077710"
 /translation="MEMNRRTLVLSQCFNTLSPDPPRTAERALSADAQNYGL
 ARLRLVAPALDEDTIRHAAVFNHLSRHHPPAGDSGISIVDSEKQITLVLSLM
 LSASPRISQSLSEALTVGKDFPRAMPALPELIANTONALAGDVSVNGIIGTAS
 SIFKFSYEVRTDALVDLTKYCLDNFAPLPEIFLTKSLSDSASGGSPILKPAF
 ESORLCCTIEYSINPDLPFEFEDHMKEMWEEFKYSSNPALTEGTEITVDVR
 AATGNCINHYLEKNEDEEPCGRIENFASVMTLLDVSKSPSRLATTEATIGLVSVT
 SVHHLRFGDVNVEICOSTIYIPVSLVDEDEETEMNYIEFTRDMEGSDVDRRTI
 ACCLTKGLATYNTQVTEVNSLEDFORLSSSPKASAWKDKDAIILVSLSKRG
 GASVSTDIDQNFQNFVLELDELORSRVNSPKASLKLTLRFHIFPFWMQE
 PELVRFLEAESNVVHVAASCIEKILLVEGARGNRYAGDLSPLFLQMTNFDAL"


```

repeat_region 12549 .15363

Query Match      80.9%  Score 17.8; DB 9; Length 127682;
Best Local Similarity 90.5%  Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATATCTACATTCGAAACGA 22
|||||
Db 60288 GTATATCTACATTCGAAACGA 60268

RESULT 4
AC129980/ 132202 bp DNA linear HNG 19-AUG-2002
LOCUS Homo sapiens chromosome 15 clone CTD-2529M12 map 15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC129980.3 GI:22297424
VERSION HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE 1 (bases 1 to 132202)
JOURNAL 2 (bases 1 to 132202)
REFERENCE Homo sapiens chromosome 15, clone CTD-2529M12
AUTHORS Unpublished
2 (bases 1 to 132202)
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamet,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,
Lu,X., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 132202)
Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamet,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 19, 2002 this sequence version replaced g1:2213335.
All repeats were identified using RepeatMasker.

```

```

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22204
Center clone name: 2529_M12
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 128: contig of 128 bp in length
* 129 228: gap of 100 bp
* 229 132202: contig of 131974 bp in length.
* Location/Qualifiers
  1. 132202
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /map="15"
    /clone="CTD-2529M12"
    /clone_lib="CTD2 Human BAC"
BASE COUNT 37938 a 27502 c 28043 g 38577 t 142 others
ORIGIN

Query Match      80.9%  Score 17.8; DB 2; Length 132202;
Best Local Similarity 90.5%  Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GTATATCTACATTCGAAACGA 22
|||||
Db 16235 GTCTATCTACATTCGAAACGA 16215

RESULT 5
AC104374 144662 bp DNA linear HNG 02-JUN-2002
LOCUS Homo sapiens chromosome 15 clone CTD-2005H15 map 15, *** SEQUENCING
DEFINITION IN PROGRESS ***, 2 ordered pieces.
ACCESSION AC104374.3 GI:21313764
VERSION HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE 1 (bases 1 to 144662)
JOURNAL 2 (bases 1 to 144662)
REFERENCE Homo sapiens chromosome 15, clone CTD-2005H15
AUTHORS Unpublished
2 (bases 1 to 144662)
Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camata,J., Campioano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Gardyna,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamet,A., Karatas,A., Kells,C., Labocque,K.,
Lamasares,R., Landers,T., Lehotcky,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPeckers,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,

```

```

repeat_region      239..538      AluY repeat: matches 75..310 of consensus"
/note="LIM4 repeat: matches 2107..2438 of consensus"
repeat_region      336..538      /note="LIM4 repeat: matches 2227..2438 of consensus"
/note="LIM4 repeat: matches 2582..3669 of consensus"
misc_feature       550..1553      complement(565..1283)
/note="match: GSS: Em:AQ383567"
repeat_region      675..1155      /note="LIM4 repeat: matches 2722..3226 of consensus"
complement(698..1442)
/note="match: GSS: Em:AQ896860"
misc_feature       /note="match(906..1442)
/note="match: GSS: Em:AQ761517"
1288..1539
/note="match: GSS: Em:AQ879547"
1290..1693
/note="match: GSS: Em:AQ826107"
1294..2055
/note="match: GSS: Em:AQ901298"
1294..1860
/note="match: GSS: Em:AQ047100"
1362..2055
/note="match: GSS: Em:AQ490007"
1362..1961
/note="match: GSS: Em:AQ585548"
1362..1761
/note="match: GSS: Em:AQ275646"
1447..1837
/note="match: GSS: Em:AQ355624"
1448..1964
/note="match: GSS: Em:AQ474420"
1449..2054
/note="match: GSS: Em:AQ474420"
1450..2017
/note="match: GSS: Em:B76221"
1451..2223
/note="match: GSS: Em:AQ738677"
1451..2150
/note="match: GSS: Em:AQ738952"
1451..1896
/note="match: GSS: Em:AQ669827"
1451..1877
/note="match: GSS: Em:AQ810603"
1451..1827
/note="match: GSS: Em:AQ718640"
1466..1877
/note="match: GSS: Em:AQ692785"
1592..2208
/note="LIM1 repeat: matches 5338..5979 of consensus
LIM1 repeat: matches 5338..5979 of consensus"
2199..2404
/note="LIM4 repeat: matches 3935..4135 of consensus
LIM4 repeat: matches 3935..4135 of consensus"
complement(2626..3066)
/note="match: GSS: Em:AQ245895"
complement(2887..3061)
/note="match: GSS: Em:B91023"
3067..3302
/note="match: GSS: Em:AQ009565"
3074..3453
/note="match: GSS: Em:AQ261089"
3074..3302
/note="match: GSS: Em:AQ059372"
3407..3448
/note="MER5B repeat: matches 1..42 of consensus"
complement(3620..4114)
/note="match: GSS: Em:AQ117010"
complement(3956..4111)
/note="match: GSS: Em:AQ304953"
4136..4803
/note="match: GSS: Em:AQ531595"

repeat_region      4636..4942      /note="AluSc repeat: matches 1..308 of consensus
AluSx repeat: matches 1..311 of consensus"
5143..5591
/note="match: GSS: Em:AQ224449"
5147..5738
/note="match: GSS: Em:B60185"
5147..5659
/note="match: GSS: Em:AQ076456"
5154..5605
/note="match: GSS: Em:AQ608022"
5204..5675
/note="match: GSS: Em:AQ712434"
5702..5976
/note="LIR16A repeat: matches 132..428 of consensus
LIR16A repeat: matches 132..428 of consensus"
6212..6878
/note="match: GSS: Em:AQ321209"
6212..6775
/note="match: GSS: Em:AQ377996"
6213..6543
/note="match: GSS: Em:AQ382827"
6216..6770
/note="match: GSS: Em:AQ344327"
6221..6541
/note="match: GSS: Em:AQ052109"
6229..6703
/note="match: GSS: Em:AQ370102"
6254..6541
/note="match: GSS: Em:AQ214603"
6256..6733
/note="match: GSS: Em:AQ454556"
6272..6541
/note="match: GSS: Em:AQ095672"
6329..6686
/note="match: GSS: Em:AQ629712"
6397..6873
/note="match: GSS: Em:AQ587044"
6406..6787
/note="match: GSS: Em:AQ587048"
6546..6888
/note="match: GSS: Em:AQ052109"
6546..6741
/note="match: GSS: Em:AQ095672"
6546..6739
/note="match: GSS: Em:AQ214603"
7281..7429
/note="MER65A repeat: matches 295..445 of consensus
MER65A repeat: matches 295..445 of consensus"
7454..7774
/note="MER65A repeat: matches 15..334 of consensus
MER65A repeat: matches 15..334 of consensus"
8040..8082
/note="MAB1 repeat: matches 34..77 of consensus
MAB1 repeat: matches 34..77 of consensus"
8437..8643
/note="LIPB1 repeat: matches 6008..6155 of consensus
LIPB1 repeat: matches 6008..6155 of consensus"
8933..9240
/note="LIR14A repeat: matches 2..37 of consensus
LIR14A repeat: matches 2..37 of consensus"
9246..10036
/note="LI repeat: matches 4060..4853 of consensus
LI repeat: matches 4060..4853 of consensus"
10042..10325
/note="LIR14A repeat: matches 1..344 of consensus
LIR14A repeat: matches 1..344 of consensus"
10326..12101
/note="HERV4 repeat: matches 1..1888 of consensus
HERV4 repeat: matches 1..1888 of consensus"
12102..12546
/note="MER9 repeat: matches 1..509 of consensus
MER9 repeat: matches 1..509 of consensus"
```

```

* 22588 23270: contig of 683 bp in length
* 23271 23370: gap of 100 bp
* 23371 24068: contig of 698 bp in length
* 24069 24168: gap of 100 bp
* 24169 24855: contig of 687 bp in length
* 24856 24955: gap of 100 bp
* 24956 25652: contig of 697 bp in length
* 25653 25752: gap of 100 bp
* 25753 26429: contig of 677 bp in length
* 26430 26529: gap of 100 bp
* 26530 27231: contig of 702 bp in length
* 27232 27331: gap of 100 bp
* 27332 28045: contig of 714 bp in length
* 28046 28145: gap of 100 bp
* 28146 28868: contig of 723 bp in length
* 28869 28968: gap of 100 bp
* 28969 29688: contig of 720 bp in length
* 29689 29788: gap of 100 bp
* 29789 30507: contig of 719 bp in length
* 30508 30607: gap of 100 bp
* 30608 31313: contig of 706 bp in length
* 31314 31413: gap of 100 bp
* 31414 32105: contig of 692 bp in length
* 32106 32205: gap of 100 bp
* 32206 32900: contig of 695 bp in length
* 32901 33000: gap of 100 bp
* 33001 33684: contig of 684 bp in length
* 33685 33784: gap of 100 bp
* 33785 34486: contig of 702 bp in length
* 34487 34586: gap of 100 bp
* 34587 35310: contig of 724 bp in length
* 35311 35410: gap of 100 bp
* 35411 36129: contig of 719 bp in length
* 36130 36229: gap of 100 bp
* 36230 36944: contig of 715 bp in length
* 36945 37044: gap of 100 bp
* 37045 37759: contig of 715 bp in length
* 37760 37859: gap of 100 bp
* 37860 38552: contig of 693 bp in length
* 38553 38652: gap of 100 bp
* 38653 39374: contig of 722 bp in length
* 39375 39474: gap of 100 bp
* 39475 40190: contig of 716 bp in length
* 40191 40290: gap of 100 bp
* 40291 41008: contig of 718 bp in length
* 41009 41108: gap of 100 bp
* 41109 41800: contig of 692 bp in length
* 41801 41900: gap of 100 bp
* 41901 42597: contig of 697 bp in length
* 42598 42697: gap of 100 bp
* 42698 43390: contig of 693 bp in length
* 43391 43490: gap of 100 bp
* 43491 44186: contig of 696 bp in length
* 44187 44286: gap of 100 bp
* 44287 44496: contig of 710 bp in length
* 44967 45066: gap of 100 bp
* 45067 45815: contig of 719 bp in length
* 45816 45915: gap of 100 bp
* 45916 46631: contig of 716 bp in length
* 46632 46731: gap of 100 bp
* 46732 47448: contig of 718 bp in length
* 47450 47549: gap of 100 bp
* 47550 48275: contig of 726 bp in length
* 48276 48375: gap of 100 bp
* 48376 49086: contig of 711 bp in length
* 49087 49186: gap of 100 bp
* 49187 49910: contig of 724 bp in length
* 49911 50010: gap of 100 bp
* 50011 50723: contig of 713 bp in length
* 50724 50823: gap of 100 bp
* 50824 51512: contig of 689 bp in length
* 51513 51612: gap of 100 bp
* 51613 52318: contig of 706 bp in length

```

```

Query Match      80.9%; Score 17.8; DB 2; Length 72579;
Best Local Similarity 90.5%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY      2  GTATCTCATTCGAAACGA 22
Db      43249  GTCTATCTCATTCGAAACGA 43269

```

```

RESULT 3
AL137861/c
LOCUS
DEFINITION
Human DNA sequence from clone RP4-813B7 on chromosome 1. Contains a
pseudogene similar to zinc finger protein ZNF186, STS and GSSs.

```

```

ACCESSION
AL137861
VERSION
AL137861.5 GI:9187172
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE
1 (bases 1 to 127682)
AUTHORS
Cobley V.
JOURNAL
Submitted (18-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk

```

On Jul 14, 2000 this sequence version replaced gi:8894204.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi; EMBL; Swi; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>

RP4-813B7 is from the library RPCI-4 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP4-813B7.

location/Qualifiers

1. 127682

FEATURES

source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/clone="RP4-813B7"

/clone_lib="RPCI-4"

1. .235

repeat_region

/note="AluX repeat: matches 75. .310 of consensus

TITLE	Journal	COMMENT	FEATURES
Direct Submission	Submitted (30-DEC-1999)	Celera Genomics, 45 West Gude Drive, Rockville, MD, USA	
		This sequence was identified as CDM:10210993 by the submitter. For more information on this record e-mail to fly@celera.com.	
		* NOTE: This is a 'working draft' sequence.	
		* This sequence will be replaced	
		* by the finished sequence as soon as it is available and	
		* the accession number will be preserved.	
source		location/Qualifiers	
		1. 20306	
		/organism="Drosophila melanogaster"	
		/db_xref="taxon:7227"	
BASE COUNT	5977 a	4332 c	4420 g 5577 t
ORIGIN			
Query Match	80.9%	Score 17.8:	DB 2: Length 20306:
Best Local Similarity	90.5%	Pred. NO. 2.8e+02:	
Matches	19;	Conservative 0;	Mismatches 2; Indels 0; Gaps 0;
Oy	2	GGATATCTACATTCGAACGA	22
Db	16836	GTATACACTGCATTCGATCGA	16856
RESULT 2	AC103875	72579 bp	DNA linear HTG 29-NOV-2001
LOCUS	AC103875		
DEFINITION	Homo sapiens chromosome 11 clone CTD-201002 map 11, LOW-PASS		
ACCESSION	AC103875		
VERSION	AC103875.1	GI:17149760	
KEYWORDS	HTG; HTGS; PHASE0.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 72579)		
TITLE	Blair, B., Linton, L., Nussbaum, C. and Lander, E.		
JOURNAL	Homo sapiens chromosome 11, clone CTD-2010D2		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 72579)		
	Blair, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barra, N., Bastien, V., Boguski, L., Bouckalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hago, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Labrecque, K., Lamacarez, R., Landers, T., Lehoczy, V., Levine, R., Liu, G., MacLean, C., MacDonald, P., Major, J., Marcus, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneses, L., Mihova, T., Mienda, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, W., Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strass, N., Subramanian, A., Talamas, J., Testaye, S., Theodores, J., Toph, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-NOV-2001)	Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT	All repeats were identified using RepeatMasker:		
	Smith, A.F.A. & Green, P. (1996-1997)		
	http://ftp.genome.washington.edu/RM/RepeatMasker.html		
	Genome Center		
	Center: Whitehead Institute/ MIT Center for Genome Research		
	Center code: WIRS		

```

Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L22053
Center clone name: 2010_D_12

** NOTE: This record contains 90 individual
** sequencing reads that have not been assembled into
** contigs. Runs of N are used to separate the reads
** and the order in which they appear is completely
** arbitrary. Low-pass sequence sampling is useful for
** identifying clones that may be gene-rich and allow
** overlap relationships among clones to be deduced.
** However, it should not be assumed that this clone
** will be sequenced to completion. In the event that
** the record is updated, the accession number will
** be preserved.

1
706 805: gap of 100 bp in length
806 1528: contig of 723 bp in length
1529 1628: gap of 100 bp in length
1629 2350: contig of 722 bp in length
2351 2450: gap of 100 bp in length
2451 3163: contig of 713 bp in length
3164 3263: gap of 100 bp in length
3264 3990: contig of 727 bp in length
3991 4090: gap of 100 bp in length
4091 4797: contig of 707 bp in length
4798 4897: gap of 100 bp in length
4898 5589: contig of 692 bp in length
5590 5689: gap of 100 bp in length
5690 6595: contig of 706 bp in length
6596 6495: gap of 100 bp in length
6496 7182: contig of 687 bp in length
7183 7282: gap of 100 bp in length
7283 7965: contig of 683 bp in length
7966 8065: gap of 100 bp in length
8066 8881: contig of 716 bp in length
8882 8881: gap of 100 bp in length
8882 9562: contig of 681 bp in length
9563 9662: gap of 100 bp in length
9663 10388: contig of 726 bp in length
10389 10488: gap of 100 bp in length
10489 11210: contig of 722 bp in length
11211 11310: gap of 100 bp in length
11311 12031: contig of 721 bp in length
12032 12131: gap of 100 bp in length
12132 12834: contig of 703 bp in length
12835 12934: gap of 100 bp in length
12935 13626: contig of 692 bp in length
13627 13726: gap of 100 bp in length
13727 14427: contig of 701 bp in length
14428 14527: gap of 100 bp in length
14528 15528: contig of 701 bp in length
15529 15328: gap of 100 bp in length
15529 16014: contig of 686 bp in length
16015 16114: gap of 100 bp in length
16115 16628: contig of 714 bp in length
16628 16928: gap of 100 bp in length
16929 17650: contig of 722 bp in length
17651 17750: gap of 100 bp in length
17751 18454: contig of 704 bp in length
18455 18554: gap of 100 bp in length
18555 19218: contig of 664 bp in length
19219 19318: gap of 100 bp in length
19319 20026: contig of 708 bp in length
20027 20126: gap of 100 bp in length
20127 20850: contig of 724 bp in length
20851 20950: gap of 100 bp in length
20951 21668: contig of 718 bp in length
21669 21768: gap of 100 bp in length
21769 22487: contig of 719 bp in length
22488 22587: gap of 100 bp in length

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 78.1344 Seconds

(without alignments)
8194.363 Million cell updates/sec

Title: US-09-477-082-30

Perfect score: 22

Sequence: 1 cgatattacattcgaacga 22

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_on:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	80.9	20306	2 AC019801	AC019801 Drosophila
2	17.8	80.9	72379	2 AC019801	AC019801 Homo sapi
3	17.8	80.9	127682	2 AL137861	AL137861 Human DNA
4	17.8	80.9	132202	2 AC0129980	AC0129980 Homo sapi
5	17.8	80.9	144662	2 AC0104374	AC0104374 Homo sapi
6	17.8	80.9	169856	3 AC012162	AC012162 Drosophila
7	17.8	80.9	173281	3 AC012161	AC012161 Drosophila
8	17.8	80.9	300994	3 AE003506	AE003506 Drosophila
9	17.2	78.2	2438	8 AY093111	AY093111 Arabidops
10	17.2	78.2	2919	8 ATH297282	AJ297282 Arabidops
11	17.2	78.2	9289	6 AX251255	AX251255 Sequence
12	17.2	78.2	11101	1 AE001191	AE001191 Treponema
13	17.2	78.2	82454	8 AC012193	AC012193 Arabidops
14	17.2	78.2	84825	8 AC006418	AC006418 Arabidops
15	17.2	78.2	95214	8 AC006526	AC006526 Arabidops
16	17.2	78.2	99492	8 ATF18021	AL163763 Arabidops
17	17.2	78.2	128739	2 AC016602	AC016602 Mus muscu
18	17.2	78.2	197662	2 AC099713	AC099713 Mus muscu
19	17.2	78.2	215126	2 AC088316	AC088316 Mus muscu
20	17.2	78.2	223580	2 AC079216	AC079216 Mus muscu
21	17.2	78.2	248201	2 AC115293	AC115293 Mus muscu
22	16.8	76.4	1501	10 AF244366	AF244366 Rattus no
23	16.8	76.4	3444	1 BAN249463	AE013765 Versinia
24	16.8	76.4	11108	1 AE013765	AE013765 Versinia
25	16.8	76.4	27559	8 SC8339	249210 S. cerevisia
26	16.8	76.4	58427	8 ATF1515	AL101023 Arabidops
27	16.8	76.4	100310	2 F17A13	AL096692 Arabidops
28	16.8	76.4	133256	2 AC127021	AC127021 Medicago
29	16.8	76.4	153382	9 HS111B22	Z98200 Human DNA s
30	16.8	76.4	157375	2 AC122095	AC122095 Rattus no
31	16.8	76.4	157903	9 AP002961	AP002961 Homo sapi
32	16.8	76.4	159635	2 AC011009	AC011009 Homo sapi
33	16.8	76.4	159705	2 AC021007	AC021007 Homo sapi
34	16.8	76.4	161687	9 AC048351	AC048351 Homo sapi
35	16.8	76.4	174499	2 AC068694	AC068694 Homo sapi
36	16.8	76.4	175179	2 AC072025	AC072025 Homo sapi
37	16.8	76.4	175977	2 AC058808	AC058808 Homo sapi
38	16.8	76.4	192155	2 AC123462	AC123462 Rattus no
39	16.8	76.4	194143	8 ATCHRIV48	AL161548 Arabidops
40	16.8	76.4	200001	8 ATCHRIV71	AL161575 Arabidops
41	16.8	76.4	204993	2 AC024447	AC024447 Homo sapi
42	16.8	76.4	205158	2 AC068980	AC068980 Homo sapi
43	16.8	76.4	256674	2 AC023594	AC023594 Homo sapi
44	16.8	76.4	288652	2 AC092942	AC092942 Homo sapi
45	16.8	76.4	313050	1 AJ414152	AJ414152 Versinia

ALIGNMENTS

```
RESULT 1
LOCUS AC019801
DEFINITION AC019801 20306 bp DNA linear HTG 03-JAN-2000
ACCESSION AC019801
VERSION AC019801.1 GI:6665096
KEYWORDS HTG; HTGS; PHASE2.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 20306)
Adams, M. and Venter, J.C.
```

```

QY      1 TAGGGATTCCGAGATTGCG 20
          |||||
Db      1016 TCGGGATTACAGACATTGCG 997

```

APPLICANT: Hoffman, Neil
APPLICANT: Hurlan, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2027 (PARA-016PRV)
CURRENT APPLICATION NUMBER: US/09/770,444
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/198,502
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 192
LENGTH: 470
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-444-192

Query Match 72.4%; Score 15.2; DB 10; Length 470;
Best Local Similarity 85.0%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTCGGAGATTGCG 20
DB 367 TATCGATTCGAGATGCG 348

RESULT 10
US-09-917-800A-861
Sequence 861, Application US/09917800A
Patent No. US20020119462A1
GENERAL INFORMATION:
APPLICANT: Mendrick, Donna
APPLICANT: Porter, Mark
APPLICANT: Johnson, Kory
APPLICANT: Castle, Arthur
APPLICANT: Elashoff, Michael
TITLE OF INVENTION: Molecular Toxicology Modeling
FILE REFERENCE: 44921-5038-US
CURRENT APPLICATION NUMBER: US/09/917,800A
CURRENT FILING DATE: 2001-07-31
PRIOR APPLICATION NUMBER: US 60/222,040
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/222,880
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: US 60/290,029
PRIOR FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: US 60/290,645
PRIOR FILING DATE: 2001-05-15
PRIOR APPLICATION NUMBER: US 60/292,336
PRIOR FILING DATE: 2001-05-22
PRIOR APPLICATION NUMBER: US 60/295,798
PRIOR FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 60/297,457
PRIOR FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: US 60/298,884
PRIOR FILING DATE: 2001-06-19
PRIOR APPLICATION NUMBER: US 60/303,459
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 1740
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 861
LENGTH: 721
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1170752
US-09-917-800A-861

Query Match 72.4%; Score 15.2; DB 10; Length 721;
Best Local Similarity 85.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGCGATTCGAGATTGCGA 21
DB 297 AGCGCTTCTGAGACTGCA 316

RESULT 11
US-09-738-626-595/c
Sequence 595, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIICHI
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentln ver. 3.0
SEQ ID NO 595
LENGTH: 1074
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-595

Query Match 72.4%; Score 15.2; DB 9; Length 1074;
Best Local Similarity 85.0%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TAGGGATTCGGAGATTGCG 20
DB 568 TAGGGACTCGAGGTTGGG 549

RESULT 12
US-10-128-714-563/c
Sequence 563, Application US/10128714
Publication No. US20030119013A1
GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Hu, Wendi
APPLICANT: Tishkoff, Daniel
APPLICANT: Zamudio, Carlos
APPLICANT: Eroskin, Alexey M
APPLICANT: Lemieux, Sebastien M
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
FILE REFERENCE: 10182-018-999
CURRENT APPLICATION NUMBER: US/10/128,714
CURRENT FILING DATE: 2002-04-23
PRIOR APPLICATION NUMBER: US 60/285,697
PRIOR FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: US 60/287,066
PRIOR FILING DATE: 2001-04-27
PRIOR APPLICATION NUMBER: US 60/295,890
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/303,899
PRIOR FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 60/316,362
PRIOR FILING DATE: 2001-08-31

```
RESULT 6
US-09-294-093B-1232/c
; Sequence 1232, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1232
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343985H1
; NAME/KEY: unsure
; LOCATION: 17, 51, 166
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-1232

Query Match      72.4%; Score 15.2; DB 10; Length 282;
Best Local Similarity 85.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AGGGGATTCGAGATTCGGA 21
      11 111 111111111 1111
DB      273 AGTGACTCGAGATTCGCGA 254

RESULT 7
US-09-294-093B-775/c
; Sequence 775, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 775
; LENGTH: 285
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343292H1
; NAME/KEY: unsure
; LOCATION: 13, 44, 47
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-775

Query Match      72.4%; Score 15.2; DB 10; Length 285;
Best Local Similarity 85.0%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AGGGGATTCGAGATTCGGA 21
      11 111 111111111 1111
```

```
DB      269 AGTGACTCGAGATTCGCGA 250

RESULT 8
US-09-770-791-820
; Sequence 820, Application US/09770791
; Patent No. US20020062014A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurdan, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2029 (PARA-018PRV)
; CURRENT APPLICATION NUMBER: US/09/770,791
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,480
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 820
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-791-820

Query Match      72.4%; Score 15.2; DB 10; Length 342;
Best Local Similarity 85.0%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AGGGGATTCGAGATTCGGA 21
      1111 111111111 1111
DB      169 AGGGATTCGAGATTCGAGA 188

RESULT 9
US-09-770-444-192/c
; Sequence 192, Application US/09770444
; Patent No. US20020023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Mathew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
```


Sequence 29, Application US/09873880
Patent No. US2002012318A1
GENERAL INFORMATION:
APPLICANT: Sewalt, Vincent
APPLICANT: Falco, S. Carl
APPLICANT: Allen, Stephen M.
TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
FILE REFERENCE: B01192 US CIP
CURRENT APPLICATION NUMBER: US/09/873,880
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/363,321
PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: 60/094,839
PRIOR FILING DATE: July 31, 1998
NUMBER OF SEQ ID NOS: 42
SOFTWARE: Microsoft Office 97
SEQ ID NO 29
LENGTH: 1362
TYPE: DNA
ORGANISM: Glycine max
US-09-873-880-29

Query Match 75.2%, Score 15.8; DB 10; Length 1362;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGATTGCGAGATTGCG 20
DB 360 AGGGATTGCGAGATTGCG 342

RESULT 3
US-10-037-270-446
Sequence 446, Application US/10037270
Publication No. US20030104529A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Vinod
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyun
APPLICANT: Chen, Ruihong
APPLICANT: Zhao, Qing A.
APPLICANT: Weinman, Tom
APPLICANT: Xue, Aidong J.
APPLICANT: Yang, Yonghong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yungqing
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Zhiwei
APPLICANT: Tillinghast, John
APPLICANT: Dirmac, Radjoe T.
TITLE OF INVENTION: No. US20030104529A1 Nucleic acids and
FILE REFERENCE: Polypeptides
FILE REFERENCE: 784CIPB
CURRENT APPLICATION NUMBER: US/10/037,270
CURRENT FILING DATE: 2002-01-04
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 1104
SOFTWARE: PC_FL_genes Version 1.0
SEQ ID NO 446
LENGTH: 1471
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (904)..(1293)
US-10-037-270-446

Query Match 73.3%; Score 15.4; DB 9; Length 1471;
Best Local Similarity 94.1%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGATTGCGAGATTGCG 19
DB 669 GAGATTGCGAGATTGCG 685
RESULT 4
US-09-961-527A-6/C
Sequence 6, Application US/09961527A
Patent No. US20020142324A1
GENERAL INFORMATION:
APPLICANT: Wang, Xun
APPLICANT: Turgeon, B. Gillian
APPLICANT: Yoder, Olen
APPLICANT: Wu, Jianguo
TITLE OF INVENTION: Fungal target genes and methods to identify those genes
FILE REFERENCE: TM0129-UT
CURRENT APPLICATION NUMBER: US/09/961,527A
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/234,673
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US 60/234,650
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 19
SEQ ID NO 6
LENGTH: 8091
TYPE: DNA
ORGANISM: Cochliobolus
US-09-961-527A-6

Query Match 73.3%; Score 15.4; DB 10; Length 8091;
Best Local Similarity 94.1%; Pred. No. 2.3e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGGATTGCGAGATTGCG 20
DB 5710 GGGATTGCGAGATTGCG 5694

RESULT 5
US-09-961-527A-1/C
Sequence 1, Application US/09961527A
Patent No. US20020142324A1
GENERAL INFORMATION:
APPLICANT: Wang, Xun
APPLICANT: Turgeon, B. Gillian
APPLICANT: Yoder, Olen
APPLICANT: Wu, Jianguo
TITLE OF INVENTION: Fungal target genes and methods to identify those genes
FILE REFERENCE: TM0129-UT
CURRENT APPLICATION NUMBER: US/09/961,527A
CURRENT FILING DATE: 2001-09-24
PRIOR APPLICATION NUMBER: US 60/234,673
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: US 60/234,650
PRIOR FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 19
SEQ ID NO 1
LENGTH: 14955
TYPE: DNA
ORGANISM: Cochliobolus
US-09-961-527A-1

Query Match 73.3%; Score 15.4; DB 10; Length 14955;
Best Local Similarity 94.1%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGGATTGCGAGATTGCG 20
DB 12555 GGGATTGCGAGATTGCG 12539

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 10.6338 Seconds
(without alignments)
3082.690 Million cell updates/sec

Title: US-09-477-082-29

Perfect score: 21
Sequence: 1 tagggagattcgagattcgga 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.8	75.2	495	US-09-873-880-7	Sequence 7, Appli
C 2	15.8	75.2	1362	US-09-873-880-29	Sequence 29, Appl
C 3	15.4	73.3	1471	US-10-037-270-446	Sequence 46, Appl
C 4	15.4	73.3	8091	US-09-861-527A-6	Sequence 6, Appli
C 5	15.4	73.3	14955	US-09-861-527A-1	Sequence 1, Appli
C 6	15.2	72.4	282	US-09-294-093B-1232	Sequence 1232, Ap
C 7	15.2	72.4	285	US-09-294-093B-775	Sequence 775, App
C 8	15.2	72.4	342	US-09-770-791-820	Sequence 820, App
C 9	15.2	72.4	470	US-09-770-444-192	Sequence 192, App
C 10	15.2	72.4	721	US-09-917-800A-881	Sequence 881, App
C 11	15.2	72.4	1074	US-09-738-626-595	Sequence 595, App
C 12	15.2	72.4	4048	US-10-128-714-563	Sequence 563, App
C 13	15.2	72.4	4048	US-10-128-714-563	Sequence 563, App
C 14	15.2	72.4	9249	US-10-224-413-67	Sequence 67, Appl
C 15	15.2	72.4	42571	US-10-224-413-3	Sequence 3, Appli
C 16	15.2	72.4	3309400	US-09-738-626-1	Sequence 1, Appli
C 17	14.8	70.5	387	US-09-878-574-1714	Sequence 1714, Ap
C 18	14.8	70.5	461	US-09-918-995-35322	Sequence 35322, A
C 19	14.6	69.5	419	US-10-056-359-1	Sequence 1, Appli

C 20	14.6	69.5	419	12	US-10-056-359-3	Sequence 3, Appli
C 21	14.6	69.5	419	12	US-10-056-360-1	Sequence 1, Appli
C 22	14.6	69.5	419	12	US-10-056-360-3	Sequence 3, Appli
C 23	14.6	69.5	436	10	US-09-878-574-4901	Sequence 4901, Ap
C 24	14.6	69.5	465	10	US-09-960-382-9325	Sequence 9325, Ap
C 25	14.6	69.5	485	9	US-09-770-961-807	Sequence 807, App
C 26	14.6	69.5	493	9	US-09-770-961-604	Sequence 604, App
C 27	14.6	69.5	1070	9	US-10-205-534-6	Sequence 6, Appli
C 28	14.6	69.5	1206	9	US-09-938-842A-2053	Sequence 2053, Ap
C 29	14.6	69.5	1210	10	US-09-887-576-477	Sequence 477, App
C 30	14.6	69.5	1210	10	US-09-887-576-577	Sequence 577, App
C 31	14.6	69.5	1302	10	US-09-815-242-7258	Sequence 7258, App
C 32	14.6	69.5	1440	10	US-09-881-732A-273	Sequence 273, App
C 33	14.6	69.5	2000	10	US-09-887-576-330	Sequence 330, App
C 34	14.6	69.5	2004	10	US-09-887-576-174	Sequence 174, App
C 35	14.6	69.5	5118	9	US-10-172-086-42	Sequence 42, Appl
C 36	14.6	69.5	6112	9	US-10-172-086-33	Sequence 33, Appl
C 37	14.6	69.5	6284	10	US-09-764-877-3643	Sequence 3643, Ap
C 38	14.6	69.5	7317	9	US-10-239-676-48	Sequence 48, Appl
C 39	14.4	68.6	479	10	US-09-864-761-14337	Sequence 14337, A
C 40	14.4	68.6	509	9	US-09-918-995-7247	Sequence 7247, Ap
C 41	14.4	68.6	972	9	US-10-128-714-7265	Sequence 7265, Ap
C 42	14.4	68.6	981	10	US-09-815-242-7388	Sequence 7388, Ap
C 43	14.4	68.6	1207	9	US-10-128-714-6265	Sequence 6265, Ap
C 44	14.4	68.6	2596	9	US-10-233-131-7	Sequence 7, Appli
C 45	14.4	68.6	2596	10	US-09-808-701-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-873-880-7/c
; Sequence 7, Application US/09873880
; Patent No. US20020123118A1
; GENERAL INFORMATION:
; APPLICANT: Sewalt, Vincent
; APPLICANT: Falco, S. Carl
; APPLICANT: Allen, Stephen M.
; TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES
; FILE REFERENCE: BBI192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (382)
; OTHER INFORMATION: n=A, C, G, or T
; NAME/KEY: unsure
; LOCATION: (454)
; OTHER INFORMATION: n=A, C, G, or T
US-09-873-880-7

Query Match 75.2%; Score 15.8; DB 10;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGAGTTGCGAGATTGCG 20
|||||
Db 353 AGGGAGTTGCGAGATTGCG 335

RESULT 2
US-09-873-880-29/c

```

:
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WordPerfect 6.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/00408
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/181,492
: FILING DATE: 13-January-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: YANKWICH, Leon R
: REGISTRATION NUMBER: 30,237
: REFERENCE/DOCKET NUMBER: TCS-203-PCT(94,664-A)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-345-9100
: TELEFAX: 617-345-9111
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 276 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical:
: ANTI-SENSE:
: FEATURE:
: NAME/KEY:
: LOCATION:
:
PCT-US95-00408-25

Query Match          67.6%; Score 14.2; DB 5; Length 276;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AGGGGATTCGAGATTGCG 20
      ||| || ||||| |||
Db      26 AGGTGAGTCGAGATTGGG 8
```

Search completed: July 8, 2003, 05:56:22
Job time : 79.0371 secs

LOCATION: 1...3363
US-09-221-017B-862

Query Match 69.5%; Score 14.6; DB 4; Length 3363;
Best Local Similarity 81.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTGCGA 21
DB 3191 TAGGCTTCGAGATTGCGA 3171

RESULT 13
US-08-181-492B-25/c
Sequence 25, Application US/08181492B
Patent No. 5552300

GENERAL INFORMATION:
APPLICANT: Makrides, Savvas C
APPLICANT: Kung, Patrick C
TITLE OF INVENTION: T Cell Antigen Receptor V Region
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: T Cell Sciences, Inc.
STREET: 115 Fourth Avenue
CITY: Needham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 02194-2725

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS 6.2
SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,492B
FILING DATE: 13-January-1994
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Yankwich, Leon R.
REGISTRATION NUMBER: 30,237
REFERENCE/DOCKET NUMBER: TCS-203-P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111

INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 276 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL:
ANTI-SENSE:
FEATURE:
NAME/KEY:
LOCATION:
US-08-181-492B-25

Query Match 67.6%; Score 14.2; DB 1; Length 276;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTCGAGATTGCG 20
DB 26 AGGTGATCGAGATTGCG 8

RESULT 14
US-08-181-492B-26
Sequence 26, Application US/08181492B

Patent No. 5552300
GENERAL INFORMATION:
APPLICANT: Makrides, Savvas C
APPLICANT: Kung, Patrick C
TITLE OF INVENTION: T Cell Antigen Receptor V Region
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: T Cell Sciences, Inc.
STREET: 115 Fourth Avenue
CITY: Needham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 02194-2725

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS 6.2
SOFTWARE: WordPerfect 6.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,492B
FILING DATE: 13-January-1994
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Yankwich, Leon R.
REGISTRATION NUMBER: 30,237
REFERENCE/DOCKET NUMBER: TCS-203-P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111

INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 276 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL:
ANTI-SENSE:
FEATURE:
NAME/KEY:
LOCATION:
US-08-181-492B-26

Query Match 67.6%; Score 14.2; DB 1; Length 276;
Best Local Similarity 84.2%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AGGGATTCGAGATTGCG 20
DB 251 AGGTGATCGAGATTGCG 269

RESULT 15
PCT-US95-00408-25/c
Sequence 25, Application PC/TUS9500408
GENERAL INFORMATION:
APPLICANT: T Cell Sciences, Inc
TITLE OF INVENTION: T Cell Antigen Receptor V Region Proteins
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: T Cell Sciences, Inc.
STREET: 1515 Fourth Avenue
CITY: Needham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 02194-2725

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC Compatible

EARLIER FILING DATE: 1995-07-19
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 419
TYPE: DNA
ORGANISM: Papilloma virus
US-09-000-266-3

Query Match 69.5% Score 14.6; DB 4: Length 419;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCGA 21
|| |||||
Db 390 TATAGATTGCGAGCTTGCA 370

RESULT 10
US-09-628-099-1
Sequence 1, Application US/09628099
Patent No. 6368632
GENERAL INFORMATION:
APPLICANT: Shamanin, Vladimir
APPLICANT: De Villiers-zur Hausen, Ethel-Michele
TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING
TITLE OF INVENTION: THEM AND FOR TREATING DISEASES CAUSED BY SUCH VIRUSES
FILE REFERENCE: 8484-0037-999
CURRENT APPLICATION NUMBER: US/09/628, 099
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 09/000, 266
PRIOR FILING DATE: 1998-10-19
PRIOR APPLICATION NUMBER: DE P 195 26 386.3
PRIOR FILING DATE: 1995-07-19
NUMBER OF SEQ ID NOS: 30
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 419
TYPE: DNA
ORGANISM: Papilloma virus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(417)
US-09-628-099-1

Query Match 69.5% Score 14.6; DB 4: Length 419;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCGA 21
|| |||||
Db 30 TATAGATTGCGAGCTTGCA 50

RESULT 11
US-09-628-099-3/C
Sequence 3, Application US/09628099
Patent No. 6368852
GENERAL INFORMATION:
APPLICANT: Shamanin, Vladimir
APPLICANT: De Villiers-zur Hausen, Ethel-Michele
APPLICANT: Zur Hausen, Harald
TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING
TITLE OF INVENTION: THEM AND FOR TREATING DISEASES CAUSED BY SUCH VIRUSES
FILE REFERENCE: 8484-0037-999
CURRENT APPLICATION NUMBER: US/09/628, 099
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 09/000, 266
PRIOR FILING DATE: 1998-10-19
PRIOR APPLICATION NUMBER: DE P 195 26 386.3
PRIOR FILING DATE: 1995-07-19
NUMBER OF SEQ ID NOS: 30

SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 419
TYPE: DNA
ORGANISM: Papilloma virus
US-09-628-099-3

Query Match 69.5% Score 14.6; DB 4: Length 419;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCGA 21
|| |||||
Db 390 TATAGATTGCGAGCTTGCA 370

RESULT 12
US-09-221-017B-862/C
Sequence 862, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221, 017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Monroy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 862:
SEQUENCE CHARACTERISTICS:
LENGTH: 3363 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

Query Match      72.4%; Score 15.2; DB 4; Length 168575;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      2  AGGGGATTGCGAGATTGCGA 21
Db      29456  AGGGGATTGCGAGATTGCGA 29475

RESULT 6
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match      72.4%; Score 15.2; DB 4; Length 4403765;
Best Local Similarity 85.0%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1  TAGGGGATTGCGAGATTGCG 20
Db      1931100  TAGAGGATTGCGAGATTGCG 1931119

RESULT 7
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
```

```

; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      72.4%; Score 15.2; DB 4; Length 4411529;
Best Local Similarity 85.0%; Pred. No. 64;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      1  TAGGGGATTGCGAGATTGCG 20
Db      1940329  TAGAGGATTGCGAGATTGCG 1940348

RESULT 8
US-09-000-266-1
; Sequence 1, Application US/09000266A
; Patent No. 6322795
; GENERAL INFORMATION:
; APPLICANT: Shamanin, Vladimir
; APPLICANT: De Villiers-Zur Hausen, Ethel-Michele
; APPLICANT: Zur Hausen, Harald
; TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING
; TITLE OF INVENTION: THEM AND FOR TREATING DISEASES CAUSED BY SUCH VIRUSES
; FILE REFERENCE: 8484-0037-999
; CURRENT APPLICATION NUMBER: US/09/000,266A
; CURRENT FILING DATE: 1998-10-19
; EARLIER APPLICATION NUMBER: PCT/DE96/01369
; EARLIER FILING DATE: 1996-07-19
; EARLIER APPLICATION NUMBER: DE P 195 26 386.3
; EARLIER FILING DATE: 1995-07-19
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Papilloma virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(417)
US-09-000-266-1

Query Match      69.5%; Score 14.6; DB 4; Length 419;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1  TAGGGGATTGCGAGATTGCGA 21
Db      30  TATAGGATTGCGAGATTGCGA 50

RESULT 9
US-09-000-266-3/C
; Sequence 3, Application US/09000266A
; Patent No. 6322795
; GENERAL INFORMATION:
; APPLICANT: Shamanin, Vladimir
; APPLICANT: De Villiers-Zur Hausen, Ethel-Michele
; APPLICANT: Zur Hausen, Harald
; TITLE OF INVENTION: PAPILLOMA VIRUSES, AGENTS FOR DETECTING
; TITLE OF INVENTION: THEM AND FOR TREATING DISEASES CAUSED BY SUCH VIRUSES
; FILE REFERENCE: 8484-0037-999
; CURRENT APPLICATION NUMBER: US/09/000,266A
; CURRENT FILING DATE: 1998-10-19
; EARLIER APPLICATION NUMBER: PCT/DE96/01369
; EARLIER FILING DATE: 1996-07-19
; EARLIER APPLICATION NUMBER: DE P 195 26 386.3
```

```
;
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-983-502-14

Query Match      84.8%; Score 17.8; DB 4; Length 2887;
Best Local Similarity 90.5%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGGATTCGGAGATTGCGA 21
Db      221 TAGGGGACTCGAGACTGCGA 241

RESULT 2
PCT-US96-10521-14
; Sequence 14, Application PC/TUS9610521
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; TITLE OF INVENTION: AND OTHER PROTEINS
; NUMBER OF SEQUENCES: 34
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US96-10521-14

Query Match      84.8%; Score 17.8; DB 5; Length 2887;
Best Local Similarity 90.5%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGGATTCGGAGATTGCGA 21
Db      221 TAGGGGACTCGAGACTGCGA 241

RESULT 3
US-09-134-246-6/c
; Sequence 6, Application US/09134246B
```

```
; Patent No. 6207377
; GENERAL INFORMATION:
; APPLICANT: Wayne, Jay
; APPLICANT: Xu, Shuang-yong
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid
; FILE REFERENCE: Thermus Replication Origins
; CURRENT APPLICATION NUMBER: US/09/134,246B
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5849
; TYPE: DNA
; ORGANISM: Thermus sp.
; US-09-134-246-6

Query Match      72.4%; Score 15.2; DB 4; Length 5849;
Best Local Similarity 85.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AGGGGATTCGGAGATTGCGA 21
Db      4909 AGGAGATTAGGAGATTGAGA 4890

RESULT 4
US-09-810-347-3
; Sequence 3, Application US/09810347
; Patent No. 6461847
; GENERAL INFORMATION:
; APPLICANT: Yi, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USBS
; FILE REFERENCE: CLO01169
; CURRENT APPLICATION NUMBER: US/09/810,347
; CURRENT FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 42571
; TYPE: DNA
; ORGANISM: Human
; US-09-810-347-3

Query Match      72.4%; Score 15.2; DB 4; Length 42571;
Best Local Similarity 85.0%; Pred. No. 1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 AGGGGATTCGGAGATTGCGA 21
Db      22240 AGGTAATCTGAGATTGCGA 22259

RESULT 5
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Bergling Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 5.03708 Seconds

(without alignments)
1278.561 Million cell updates/sec

Title: US-09-477-082-29

Sequence: 1 taggggattcgagattcgca 21

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_NA:*
1: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/prodata/1/lna/5A.COMB.seq:*
4: /cgn2_6/prodata/1/lna/5B.COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/prodata/1/lna/Backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	2887	4	US-08-983-502-14
2	17.8	84.8	2887	4	PCT-US96-10521-14
3	15.2	72.4	5849	4	US-09-134-245-6
4	15.2	72.4	42571	4	US-09-810-347-3
5	15.2	72.4	168575	4	US-09-426-290-1
6	15.2	72.4	4403765	4	US-09-103-840A-2
7	15.2	72.4	4411529	4	US-09-103-840A-1
8	14.6	69.5	419	4	US-09-000-265-1
9	14.6	69.5	419	4	US-09-000-265-3
10	14.6	69.5	419	4	US-09-028-099-1
11	14.6	69.5	419	4	US-09-028-099-3
12	14.6	69.5	3363	4	US-09-221-017B-862
13	14.2	67.6	276	1	US-08-181-492B-25
14	14.2	67.6	276	1	US-08-181-492B-25
15	14.2	67.6	276	5	PCT-US95-00408-25
16	14.2	67.6	1188	4	US-09-064-693A-17
17	14.2	67.6	1312	4	US-09-193-792-20
18	14.2	67.6	2021	4	US-09-193-792-2
19	14.2	67.6	2284	4	US-09-193-792-1
20	14.2	67.6	4086	4	US-08-313-181-1
21	14.2	67.6	6641	4	US-09-064-693A-25
22	14.2	67.6	7445	4	US-09-178-973B-8
23	14.2	67.6	7445	4	US-09-419-568B-8
24	14.2	67.6	7445	4	US-09-354-243B-8
25	14.2	67.6	24417	2	US-08-846-762-1
26	14.2	67.6	33529	4	US-09-144-085-3
27	14.2	67.6	4403765	4	US-09-103-840A-2

C 28	14.2	67.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
C 29	13.8	65.7	276	4	US-08-991-789A-32	Sequence 32, Appl1
C 30	13.8	65.7	276	4	US-09-062-451-32	Sequence 32, Appl1
C 31	13.8	65.7	276	4	US-09-598-326-32	Sequence 32, Appl1
C 32	13.8	65.7	420	1	US-08-470-179-162	Sequence 162, Appl1
C 33	13.8	65.7	460	2	US-08-487-727A-1	Sequence 1, Appl1
C 34	13.8	65.7	1484	2	US-08-185-828A-22	Sequence 22, Appl1
C 35	13.8	65.7	1532	2	US-08-185-828A-11	Sequence 11, Appl1
C 36	13.8	65.7	2288	4	US-08-935-433-1	Sequence 1, Appl1
C 37	13.8	65.7	2288	4	US-09-553-132-1	Sequence 1, Appl1
C 38	13.8	65.7	4137	4	US-09-489-964-2	Sequence 2, Appl1
C 39	13.8	65.7	6914	1	US-08-920-827-22	Sequence 22, Appl1
C 40	13.8	65.7	6914	1	US-08-920-827-22	Sequence 22, Appl1
C 41	13.8	65.7	6914	1	US-08-921-177-22	Sequence 22, Appl1
C 42	13.8	65.7	6914	1	US-08-362-577C-22	Sequence 22, Appl1
C 43	13.8	65.7	6914	2	US-08-920-828-22	Sequence 22, Appl1
C 44	13.8	65.7	8912	4	US-08-469-260A-11	Sequence 11, Appl1
C 45	13.8	65.7	9143	2	US-08-639-857-32	Sequence 32, Appl1

ALIGNMENTS

RESULT 1
US-08-983-502-14
; Sequence 14, Application US/08983502
; Patent No. 6399327
; GENERAL INFORMATION:
; APPLICANT: David WALLACH
; APPLICANT: Mark P. BOLDIN
; APPLICANT: Tanya M. GONCHAROV
; APPLICANT: Yury V. GOLTSSEV
; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
; TITLE OF INVENTION: AND OTHER PROTEINS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Broadway and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/983,502
; FILING DATE: 16-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10521
; FILING DATE: 14-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,615
; FILING DATE: 16-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 114,986
; FILING DATE: 17-AUG-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 115,319
; FILING DATE: 14-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 116,588
; FILING DATE: 27-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 117,932
; FILING DATE: 16-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: WALLACH-19
; TELECOMMUNICATION INFORMATION:


```

source
1.722
/organism="Pristionchus pacificus"
/strain="var. California"
/db_xref="taxon:54126"
/clone_lib="Pristionchus pacificus BAC ends"
BASE COUNT      222 a      197 c      146 g      157 t
ORIGIN

Query Match      78.1%; Score 16.4; DB 17; Length 722;
Best Local Similarity 94.4%; Pred. No. 2.2e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 GGGGATTGGAGATTGCG 20
        ||||||||||||||||
Db      230 GGGGATTGGAGATTGTG 213

RESULT 14
LOCUS      B0135457
DEFINITION NF006E01ECF1004 Elicited cell culture Medicago truncatula cDNA
ACCESSION  B0135457
VERSION    B0135457
KEYWORDS   EST.
SOURCE     barrel medic.
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
            Medicago.
REFERENCE  1 (bases 1 to 936)
AUTHORS   Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
            Flores, H.R., Imanu, J.T., Weller, J.W. and May, G.D.
TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
            Center for Medicago Genomics Research
JOURNAL    Unpublished (2000)
COMMENT    Contact: Dixon RA
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7302
            Fax: 580 221 7380
            Email: radixon@noble.org
            Insert Length: 936 Std Error: 0.00
            Plate: 005 row: E column: 01
            Seq primer: TCACACAGCAAGAACGCTATGAC.
FEATURES
Source
1.936
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF006E01EC"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
            days. Cells were induced six days after subculture"
/note="Vector: Lambda Zap; Cells were induced with yeast
            cell wall extracts equivalent to 50ug/ml glucose in the
            final concentration. Samples were taken at 0.5, 1, 12 and
            24 hours after induction. Equal amounts of RNA from each
            time point were pooled and used for mRNA isolation."
BASE COUNT      272 a      230 c      160 g      274 t
ORIGIN

Query Match      78.1%; Score 16.4; DB 14; Length 936;
Best Local Similarity 94.4%; Pred. No. 2.3e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GGGATTGGAGATTGCGA 21
        ||||||||||||||||
Db      594 GGGATTGGAGATTGAGA 577

```

```

RESULT 15
LOCUS      CNS00VWG
DEFINITION Arabidopsis thaliana genome survey sequence 17 end of BAC T10B23 of
            TBMU library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION  AL092470
VERSION    AL092470.1 GI:5293624
KEYWORDS   GSS.
SOURCE     Arabidopsis thaliana.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 280)
AUTHORS   Salanoubat, M., Cholsne, N., Attiguenave, F., Brotlier, P., Wincker, P.,
            Samsou, D., Saurin, W., Weissenbach, J. and Queller, F.
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 280)
AUTHORS   Direct Submission
TITLE     Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES
Source
1.280
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="T10B23"
/clone_lib="TBMU"
/note="end : 17"
BASE COUNT      87 a      39 c      43 g      111 t
ORIGIN

Query Match      77.1%; Score 16.2; DB 17; Length 280;
Best Local Similarity 85.7%; Pred. No. 2.2e+03;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TAGGGATTGGAGATTGCGA 21
        ||||||||||||||||
Db      26 TAGGGATTAGAGATTGCGA 46

```

Search completed: July 8, 2003, 05:51:17
Job time : 163.34 secs

QY 3 GGGGATTCGAGATTGCG 20
 |||||
 Db 234 GGGGATTCGAGATTGCG 217

RESULT 11
 BH220787 441 bp DNA linear GSS 08-NOV-2001
 LOCUS 1006097A02.1EL_X1 1006 - Rescuemu Grid G Zea mays genomic, DNA
 DEFINITION sequence.
 BH220787
 ACCESSION BH220787.1 GI:16815046
 VERSION
 KEYWORDS GSS.
 SOURCE
 ORGANISM Zea mays.
 Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 441)
 Walbot,V.
 REFERENCE 1006097A02.1EL_X1 1006 - Rescuemu Grid G Zea mays genomic, DNA
 TITLE Maize genomic sequences found using engineered Rescuemu transposon
 JOURNAL Unpublished (2001)
 COMMENT Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Very probable ligation site of ends cut by single endonuclease.
 Reverse complemented post-ligation sequence from source sequence.
 Plate: 1006097 Row: 29
 Class: transposon-tagged.
 Location/Qualifiers
 1..441
 /organism="Zea mays"
 /cultivar="mixed background W23/A188/B73"
 /db_xref="taxon:4577"
 /clone_id="1006 - Rescuemu Grid G"
 /issue_type="leaf"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: leaf; Vector: Rescuemu (engineered from
 Bluescript backbone); Site_1: BamHI; Site_2: BglII;
 Rescuemu is a 4.9 kb, modified maize Mu transposon
 designed to allow plasmid rescue from total genomic DNA.
 Mu elements insert preferentially into transcription
 units. For more information on Rescuemu, go to the web
 site 'www.zmdb.iastate.edu' and follow the links for
 'Rescuemu.' Grid G was grown at Stanford in 2000. DNA was
 extracted from leaf punches, double digested using BamHI
 and BglII, and ligated to form circular plasmids. DH10B
 cells were transformed and then screened on LB plates with
 ampicillin."

BASE COUNT 67 a 162 c 143 g 69 t
 ORIGIN

Query Match 78.1%; Score 16.4; DB 17; Length 441;
 Best Local Similarity 94.4%; Pred. No. 2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGGATTCGAGATTGCG 20
 |||||
 Db 80 GGGGATTCGAGATTGCG 97

RESULT 12
 AA390970 598 bp mRNA linear EST 19-APR-2001
 LOCUS LD09736.5prtime LD Drosophila melanogaster embryo Bluescript
 DEFINITION Drosophila melanogaster cDNA clone LD09736 5prtime, mRNA sequence.
 ACCESSION AA390970
 VERSION AA390970.1 GI:2043964

KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 598)
 Harvey,D., Broksstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
 Lewis,S. and Rubin,G.M.
 BDGP/HMT Drosophila EST project
 Unpublished (2001)
 JOURNAL
 COMMENT Contact: Stapleton, M.
 BDGP
 Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: http://www.fruitfly.org/EST_estfruitfly.berkeley.edu
 Plate: 97 Row: C Column: 12
 High quality sequence stop: 469.
 Location/Qualifiers
 1..598
 /organism="Drosophila melanogaster"
 /db_xref="BDGP:EST:BDG1009023"
 /db_xref="taxon:7227"
 /clone="LD09736"
 /clone_id="LD Drosophila melanogaster embryo Bluescript"
 /sex="male and female"
 /dev_stage="0 to 24 hours mixed stage embryonic"
 /lab_host="SOLR"
 /note="Organ: embryo; Vector: Bluescript SK; Site_1: EcoRI
 ; Site_2: XhoI; Constructed using StrataGene ZAP-cDNA
 synthesis kit. Oligo dt-primed and directionally cloned at
 EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT 199 a 105 c 135 g 159 t
 ORIGIN

Query Match 78.1%; Score 16.4; DB 9; Length 598;
 Best Local Similarity 94.4%; Pred. No. 2.1e-03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTG 18
 |||||
 Db 572 TAGGGATTCGAGAAATG 589

RESULT 13
 BH823627 722 bp DNA linear GSS 20-MAY-2002
 LOCUS BACP2-606.Y Pristionchus pacificus BAC ends Pristionchus pacificus
 DEFINITION genomic DNA sequence.
 BH823627
 ACCESSION BH823627.1 GI:21001370
 VERSION
 KEYWORDS GSS.
 SOURCE Pristionchus pacificus.
 ORGANISM Pristionchus pacificus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
 Neodiplogasteridae; Pristionchus.
 1 (bases 1 to 722)
 Sriivasan,J., Shiz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz,
 G., Witte,H., Keller,H., Kipping,I., Pires da Silva,A., Jesse,T.,
 Millare,J., de Both,M., Schuster,S.C. and Sommer,R.J.
 A BAC-based genetic linkage map of the nematode Pristionchus
 pacificus
 Unpublished (2002)
 JOURNAL
 COMMENT Contact: Sommer RJ
 Evolutionary Biology
 Max-Planck-Institute for Developmental Biology
 Spemannstr. 37-39, Tuebingen D-72076, Germany
 Tel: 00497071601371
 Fax: 00497071601498
 Email: raif.sommer@tuebingen.mpg.de
 Class: BAC ends.
 Location/Qualifiers

BASE COUNT	68 a	91 c	41 g	97 t
ORIGIN	/sex="male" /tissue_type="aorta and vein" /dex_stage="adult" /lab_host="DH10B" /note="Site_1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGATTCGAGGTATTAATAATATCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from LambdaBada			
Query Match	Best Local Similarity	94.4%	Score 16.4; DB 10; Length 297;	
Matches	17; Conservative	0; Mismatches	1; Indels	0; Gaps 0;
Qy	1 TAGGGGATTCGAGATTG 18			
Db	96 TAGGGGTTTCGAGATTG 79			
RESULT 10				
BH826660/c				
LOCUS	BH826660	377 bp	DNA	linear
DEFINITION	BACPE24-423.2 Pristionchus pacificus BAC ends Pristionchus			
ACCESSION	Pristionchus genomic, DNA sequence.			
VERSION	BH826660			
KEYWORDS	BH826660.1 GI:21006923			
SOURCE	GSS.			
ORGANISM	Pristionchus pacificus.			
REFERENCE	Pristionchus pacificus.			
AUTHORS	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.			
JOURNAL	1 (bases 1 to 377)			
COMMENT	Srinivasan,U., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Radatz,G., Mite,H., Keller,H., Kipping,T., Pires da Silva,A., Jesse,T., Millaere,T., de Both,M., Schuster,S.C. and Sommer,R.J.,			
TITLE	A BAC-based genetic linkage map of the nematode Pristionchus pacificus			
UNPUBLISHED	(2002)			
CONTACT	Sommer RJ			
EVOLUTIONARY BIOLOGY	Max-Planck-Institute for Developmental Biology			
SEMANSTR	37-39 Tuebingen D-72076, Germany			
TEL	00497071601371			
FAX	00497071601498			
EMAIL	ralf.sommer@uebingen.mpg.de			
CLASS	BAC ends.			
LOCATION/QUALIFIERS	1. 377			
ORGANISM="Pristionchus pacificus"	/strain="var. Californa"			
/db_xref="taxon:54126"	/clone_lib="Pristionchus pacificus BAC ends"			
BASE COUNT	99 a	129 c	66 g	83 t
ORIGIN				
Query Match	Best Local Similarity	94.4%	Score 16.4; DB 17; Length 377;	
Matches	17; Conservative	0; Mismatches	1; Indels	0; Gaps 0;

Db 511 TAGGGATTCTGAGATTGC 529

RESULT 4
BQ226975/c 935 bp mRNA linear EST 02-MAY-2002

LOCUS AGENCOURT_7561218 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6054859
5', mRNA sequence.

ACCESSION BQ226975
VERSION BQ226975.1 GI:20408375
KEYWORDS EST.
SOURCE human.

ORGANISM

Human sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DMP

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13314 row: c column: 20
High quality sequence stop: 512.
Location/Qualifiers

FEATURES
source

1. .935

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6054859"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 225 a 278 c 273 g 159 t

ORIGIN

Query Match 80.0%; Score 16.8; DB 14; Length 935;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGCGAGATTGCGA 21

Db 797 AGGGGATTCGCGAGATTGCGA 778

RESULT 5
BQ899695 939 bp mRNA linear EST 16-AUG-2002

LOCUS AGENCOURT_8750911 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6336113
5', mRNA sequence.

ACCESSION BQ899695
VERSION BQ899695.1 GI:22291697
KEYWORDS EST.
SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Macconochie, Ph.D. and Nancy L. Freeman,

Ph.D.
CDNA Library Preparation: Resgen, Invitrogen Corp
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13799 row: j column: 18
High quality sequence stop: 634.
Location/Qualifiers

FEATURES
source

1. .939

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6336113"
/clone_lib="NIH_MGC_130"
/lab_host="DH10B (phage-resistant)"
/note="Organ: oocytes; Vector: pCMV-SPORT6.1.ccdB;
Site_1: EcoRV; Site_2: NotI; Cloned unidirectionally.
Primer: oligo dt. Average insert size 1.95 kb.
Constructed by Resgen, Invitrogen Corp. Note: this is a
NIH_MGC Library."

BASE COUNT 250 a 197 c 281 g 210 t 1 others

ORIGIN

Query Match 80.0%; Score 16.8; DB 14; Length 939;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGCGAGATTGCGA 21

Db 793 AGGGGATTGCGAGATTGCGA 812

RESULT 6
BQ672783/c 999 bp mRNA linear EST 15-JUL-2002

LOCUS AGENCOURT_8191724 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6257092
5', mRNA sequence.

ACCESSION BQ672783
VERSION BQ672783.1 GI:21783617
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE Unpublished (1999)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM2410 row: n column: 05
High quality sequence stop: 564.
Location/Qualifiers

FEATURES
source

1. .999

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6257092"
/clone_lib="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Source

1. 421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686E18155"
/clone_1lb="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pT7Blue2; Site_1: SfiIA; Site_2: SfiIB;
CDNA-collection"

BASE COUNT 95 a 89 c 113 g 124 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 421;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTCGA 21
||||| ||||| ||||| |||||
DB 284 TAGGGACTCGAGACTCGCA 304

RESULT 2 522 bp mRNA linear EST 14-NOV-2001
AL600571
LOCUS DKFZP313F1234 .1 313 (synonym: hlcc3) Homo sapiens CDNA clone
DEFINITION DKFZP313F1234 5', mRNA sequence.
ACCESSION AL600571
VERSION AL600571.1 GI:15164077
KEYWORDS EST.

SOURCE

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 522)
Ansoerge, W., Winkner, U., Mewes, W., Well, B. and Wiemann, S.
EST (Ansoerge, W., Winkner, U., Mewes, H.W., Well, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansoerge W

MPIS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the CDNA sequencing consortium of the
German Genome Project.
No SI sequence available.
This clone (DKFZP313F1234) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

Source

1. 522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP313F1234"
/clone_1lb="313 (synonym: hlcc3)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pT7Blue2; Site_1: SfiIA; Site_2: SfiIB;
CDNA-collection"

BASE COUNT 120 a 116 c 133 g 149 t 4 others

ORIGIN

Query Match 84.8%; Score 17.8; DB 9; Length 522;
Best Local Similarity 90.5%; Pred. No. 4.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTCGA 21
||||| ||||| ||||| |||||

DB 303 TAGGGACTCGAGACTCGCA 323

RESULT 3 697 bp DNA linear GSS 20-FEB-2001
A2821487
LOCUS 2M0094H14F Mouse 10kb plasmid UGCG1M library Mus musculus genomic
DEFINITION clone UGCG2M0094H14 F. DNA sequence.
ACCESSION A2821487
VERSION A2821487.1 GI:12991395
KEYWORDS GSS.
SOURCE mouse mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 697)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)

COMMENT

Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0094 row: H column: 14
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 697.

FEATURES

Location/Qualifiers

1. 697
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG2M0094H14"
/clone_1lb="Mouse 10kb plasmid UGCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/note="Vector: PMD42ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g1473214[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 184 a 189 c 157 g 167 t

ORIGIN

Query Match 82.9%; Score 17.4; DB 17; Length 697;
Best Local Similarity 94.7%; Pred. No. 7.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTCG 19
||||| ||||| ||||| |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 : Search time 158.34 Seconds
(without alignments)
2147.949 Million cell updates/sec

Title: US-09-477-082-29

Perfect score: 21
Sequence: 1 tagggagtcgagatgcga 21

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_hnv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	421	9	AL702116 DKF2P686E
2	17.8	84.8	522	9	AL600571 DKF2P313F
3	17.4	82.9	697	17	A6821487 2M0094H14
4	16.8	80.0	935	14	BQ226975 AGENCOURT
5	16.8	80.0	939	14	BQ899695 AGENCOURT
6	16.8	80.0	999	14	BQ672783 AGENCOURT

7	16.8	80.0	1261	12	BG425308
8	16.4	78.1	233	10	BB171058
9	16.4	78.1	297	10	BB214821
10	16.4	78.1	377	17	BB826560
11	16.4	78.1	441	17	BB220787
12	16.4	78.1	598	9	AA390970
13	16.4	78.1	722	17	BB823627
14	16.4	78.1	936	14	BQ135457
15	16.2	77.1	280	17	CNSO0YMG
16	16.2	77.1	404	10	AM284750
17	16.2	77.1	466	9	A1947129
18	16.2	77.1	466	12	BG37945
19	16.2	77.1	511	10	AM078425
20	16.2	77.1	519	10	AM058858
21	16.2	77.1	519	17	DR16H14S
22	16.2	77.1	570	13	BM323271
23	16.2	77.1	613	17	DR16G17S
24	16.2	77.1	649	17	A2321056
25	16.2	77.1	673	14	BQ910990
26	16.2	77.1	738	14	BQ971522
27	16.2	77.1	874	12	BF260277
28	16.2	77.1	924	13	BM459635
29	16.2	77.1	1085	17	AG043322
30	16.2	77.1	1422	17	AG103948
31	16.2	77.1	502	14	BQ104565
32	16.2	76.2	1101	17	CNS016VT
33	15.8	75.2	158	10	BE094710
34	15.8	75.2	177	12	BE845182
35	15.8	75.2	218	9	AV293343
36	15.8	75.2	229	10	AM899260
37	15.8	75.2	324	10	BB102604
38	15.8	75.2	365	14	BB017102
39	15.8	75.2	375	12	BF777699
40	15.8	75.2	422	10	AV829200
41	15.8	75.2	470	17	BB194313
42	15.8	75.2	493	9	AA271049
43	15.8	75.2	493	12	BF056261
44	15.8	75.2	521	14	BQ339223
45	15.8	75.2	533	9	A1774687

ALIGNMENTS

RESULT 1
AL702116
LOCUS
DEFINITION
DKF2P686E18155_r1 686 (synonym: hlc03) Homo sapiens cDNA clone
ACCESSION
AL702116
VERSION
AL702116.1 GI:19685471
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
EST.
REFERENCE
1 (bases 1 to 421)
AUTHORS
EST (Bloecker, H., Boecker, M., Brandt, P., Mewes, H.W., Well, B. and Wiemann, S.)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Bloecker H

Am Kioferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ): Email s.wiemann@dkfz-heidelberg.de;
Breunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No si sequence available.
This clone (DKF2P686E18155) is available at the RZPD in Berlin.

Query Match 80.0%; Score 16.8; DB 24; length 811;
 Best Local Similarity 90.0%; Pred. No. 1e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTGCGAGATGCGA 21
 |||||
 DB 247 AGGGGATTGCGAGATGCGA 228

Search completed: July 8, 2003, 03:11:11
 Job time : 21.9513 secs

PS Claim 1; SEQ ID NO 767; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX

XX Sequence 8951 BP; 1965 A; 548 C; 2538 G; 3900 T; 0 other;

SO

Query Match 82.9%; Score 17.4; DB 24; Length 8951;

Best Local Similarity 94.7%; Pred. No. 58;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGGATCGAGATTGCGA 21

DB 5581 GGGGTTGCGAGATTGCGA 5599

RESULT 12

ABQ22580

ID ABQ22580 standard; DNA; 540 BP.

XX

AC ABQ22580;

XX

DT 12-JUL-2002 (first entry)

XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 9171.

XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX

OS Homo sapiens.

XX

PN WO200218632-A2.

XX

PD 07-MAR-2002.

XX

PF 01-SEP-2001; 2001WO-EP10074.

XX

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX

PA (EPIG-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX

DR WPI; 2002-371829/40.

XX

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

XX

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

PS

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the

CC methylation status of many C residues to be determined simultaneously.

CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the

CC method for determining the degree of cytosine methylation described in

CC the disclosure of the invention.

XX

XX Sequence 540 BP; 86 A; 57 C; 200 G; 197 T; 0 other;

SO

Query Match 80.0%; Score 16.8; DB 24; Length 540;

Best Local Similarity 90.0%; Pred. No. 1e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTCGAGATTGCG 20

DB 457 TTGGGATTCGAGATTGCG 476

RESULT 13

ABQ22581/c

ID ABQ22581 standard; DNA; 540 BP.

XX

AC ABQ22581;

XX

DT 12-JUL-2002 (first entry)

XX

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 9172.

XX

KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX

OS Homo sapiens.

XX

PN WO200218632-A2.

XX

PD 07-MAR-2002.

XX

PF 01-SEP-2001; 2001WO-EP10074.

XX

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX

PA (EPIG-) EPIGENOMICS AG.

XX

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX

DR WPI; 2002-371829/40.

XX

PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA

XX

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

PS

XX This invention describes a novel method for determining the degree of

CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert

CC cytosine (C) but not methylated C, to uracil, then part of the genomic

CC DNA that contains the target C is amplified to form a labeled amplicon.

CC The amplicon is hybridised to two classes, each with at least one

CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

CC and the degree of hybridisation to both classes is determined from the

CC label on the amplicon. From the ratio of labels hybridised to the two

CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of

CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory

CC systems etc., particularly by detecting mutations or single nucleotide

CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

CC types and for investigating cell differentiation. The method allows the

PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241122.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM:
 XX
 DR WPI: 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and

PT metastasis -
 XX
 PS Disclosure; SEQ ID NO 32029; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 45017 BP; 12919 A; 10037 C; 9566 G; 12495 T; 0 other;
 Query Match 84.8%; Score 17.8; DB 22; Length 45017;
 Best Local Similarity 90.5%; Pred. No. 40;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TAGGGATTGGAGATTGCCA 21
 Db 25422 TAGGGACTCGAGACTGCCA 25402
 RESULT 11
 ID ABL32794 standard; DNA; 8951 BP.
 XX
 AC ABL32794;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 767.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 KW
 XX
 OS Homo sapiens.
 XX
 OS
 PN W0200200928-A2.
 XX
 PF 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI: 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251956.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259578.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
XX
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
XX Disclosure; SEQ ID NO 32028; 3071bp + Sequence Listing; English.
XX
XX
XX AA54951 to AA64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AA62170 to AA61921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AA64703
CC to AA67694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AA54942 to AA54950 and AA62169
CC represent sequences used in the exemplification of the present invention.
XX
XX
XX Sequence 43938 BP; 12772 A; 9656 C; 9206 G; 12304 T; 0 other;
SQ
Query Match 84.8%; Score 17.6; DB 22; Length 43338;
Best Local Similarity 90.5%; Freq. No. 40;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TAGGGATTGCGAGATTGCGA 21
||||| ||||||| |||||
DB 24343 TAGGGACTCGAGACTGCGA 24323
RESULT 10
AAK7217/C
ID AAK7217 standard; DNA: 45017 BP.
XX
XX AAK7217;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/hematopoietic antigen genomic sequence SEQ ID NO:32029.
DE Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200157182-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01354.
PF

XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216680.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.

RESULT 9
AAK7216/C
ID AAK77216 standard: DNA; 43338 BP.
XX
AC AAK77216;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32028.
XX
KW Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
PD
XX
PD 09-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 14-JUL-2000; 2000US-0217496.
XX 26-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226686.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0228287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229512.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0233080.
XX 08-SEP-2000; 2000US-0233081.
XX 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 20-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246552.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249247.
PR 17-NOV-2000; 2000US-0249248.
PR 17-NOV-2000; 2000US-0249249.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

FT /product= "Human caspase-8"
 XX WO200129232-A2.
 XX
 XX 26-APR-2001.
 XX
 XX 19-OCT-2000; 2000WO-US28941.
 XX
 XX 20-OCT-1999; 99US-0160559.
 XX 14-AUG-2000; 2000US-0225564.
 XX
 XX (SCIO-) SCIOS INC.
 XX
 XX Cordell B, Li Y;
 XX
 XX WPI: 2001-290920/30.
 XX P-PSDB; AAE00605.
 XX
 XX Novel fusion polypeptide comprising first and second caspase subunit
 PT separated by cleavage site not associated in nature with caspase
 PT subunit, useful for cloning gene encoding enzymes involved in
 PT proteolytic cleavage -
 XX
 XX Disclosure; Fig 15; 116pp; English.
 XX
 CC The present sequence is a DNA encoding human Caspase-8 also known as
 CC MACH, FICE and Mch5. Caspases are a family of cysteine proteases, that
 CC participate in the initiation and execution of apoptosis. Caspases exist
 CC as pro-enzymes, activated by cleavage into a large and small subunit,
 CC occurring after specific aspartic acid residues within the pro-enzyme
 CC sequence. The present invention relates to a method for functional
 CC cloning of genes encoding proteins or enzymes involved in proteolytic
 CC cleavage. The invention is based on the use of caspase expression
 CC cassettes comprising the coding sequence of a proteolytic cleavage site
 CC flanked by sequences encoding two caspase subunits. A fusion polypeptide
 CC comprising a first and a second caspase subunit, separated by a cleavage
 CC site not associated in nature, is useful for cloning gene encoding
 CC enzymes involved in proteolytic cleavage. An expression cassette
 CC containing fusion polypeptide is used to identify a mutant cell line
 CC deficient in an enzyme of interest and is also useful for diagnosis and
 CC suppression of proliferation or metastases of a tumour cell characterised
 CC by overexpression of a polypeptide (e.g. Cathepsin B or urokinase,
 CC selectively expressed in the tumour cells). DNA encoding fusion
 CC polypeptide is used in gene therapy.
 CC
 XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;
 XX
 XX Query Match 84.8%; Score 17.8; DB 22; Length 2887;
 XX Best Local Similarity 90.5%; Pred. No. 36;
 XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 1 TAGGGATTGGAGATTGCGA 21
 XX ||||||| ||||||| |||||
 DB 221 TAGGGACTCGAGACTGCGA 241
 XX
 XX RESULT 8
 XX ABR4723
 XX ID ABR4723 standard; CDNA; 2887 BP.
 XX AC ABR4723;
 XX
 XX 14-AUG-2002 (first entry)
 XX
 XX Human cDNA differentially expressed in granulocytic cells #1294.
 XX
 XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
 XX viral infection; parasitic infection; protozoal infection;
 XX fungal infection; sterile inflammatory disease; psoriasis;
 XX rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 XX cardiac reperfusion injury; renal reperfusion injury; ARDS;
 XX adult respiratory distress syndrome; inflammatory bowel disease;
 XX Crohn's disease; ulcerative colitis; periodontal disease;

KM granulocyte activation; chronic inflammation; allergy.
 XX
 XX Homo sapiens.
 XX
 XX WO200228999-A2.
 XX
 XX 11-APR-2002.
 XX
 XX 03-OCT-2001; 2001WO-US30821.
 XX
 XX 03-OCT-2000; 2000US-237189P.
 XX
 XX (GENE-) GENE LOGIC INC.
 XX
 XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 XX
 XX WPI: 2002-435328/46.
 XX
 XX Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 XX Claim 1; SEQ ID NO 1294; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) Gs by contacting GC with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating GCA; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease; also bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;
 XX
 XX Query Match 84.8%; Score 17.8; DB 24; Length 2887;
 XX Best Local Similarity 90.5%; Pred. No. 36;
 XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 QY 1 TAGGGATTGGAGATTGCGA 21
 XX ||||||| ||||||| |||||
 DB 221 TAGGGACTCGAGACTGCGA 241
 XX

```

XX Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
PT prognosing cancer, comprises detecting a modification of genomic DNA
PT comprising the CASP8 gene that results in inactivation of the gene
XX
PS Claim 21, Page 97-98; 107pp; English.
XX
CC Inactivation of a caspase-8 (CASP8) gene is determined by detecting a
CC modification of CASP8 genomic DNA that results in inactivation of the
CC gene. CASP8, a cysteine protease, is part of the death induction signaling
CC complex (DISC), associated with the Fas receptor. CASP8 is inactivated in
CC cancers, and plays a role of a tumor suppressor gene. The CASP8 promoter
CC region sequences, in particular Region 1 and Region 2, are crucial to the
CC design and execution of the genomic methylation PCR analysis of CASP8
CC gene inactivation. Methylation PCR can be used to examine even minute
CC amounts of patient material to demonstrate whether the CASP8 gene
CC expresses an mRNA and protein product. The promoter Region 1 sequence is
CC located upstream (5') to exon 1, which is the alternatively-spliced 5'
CC untranslated region (UTR) that is less commonly used. The promoter
CC Region 2 sequence is located downstream (3') of exon 1 and upstream of
CC exon 2, which is the more commonly used 5' UTR. The CASP8 gene has been
CC localized to human chromosome 2q33-34. The methods are used to diagnose
CC or prognose cancer. Cancer is treated by administering a vector that
CC expresses a gene encoding functional CASP8 in cells. The cancer that
CC is diagnosed or treated is a tumour in which a myc gene is amplified,
CC such as a neuroblastoma. Aggressive neuroblastoma, juvenile neuroblastoma
CC (preferred), small-cell lung carcinoma, non-small-cell lung carcinoma,
CC colorectal carcinoma, or uterine cervical carcinoma can be diagnosed with
CC the new method. A kit for screening for a compound that induces
CC death-receptor-mediated apoptosis in cells containing an inactivated
CC CASP8 gene is also provided.
XX
SQ Sequence 753 BP; 140 A; 204 C; 171 G; 238 T; 0 other:
XX
Query Match      84.8%; Score 17.8; DB 21; Length 753;
Best Local Similarity 90.5%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TAGGGGATTGGAGATTGCGA 21
    ||||||| ||||||| |||||
Db 90 TAGGGGACTCGAGACTGCGA 110
XX
RESULT 6
AAT61405
ID AAT61405 standard; cDNA; 2887 BP.
XX
AC AAT61405;
XX
DT 29-OCT-1997 (first entry)
XX
DE MACH isoform alpha coding sequence.
XX
MACH; MORT-1 binding protein; mediator of receptor toxicity; cell death;
KM antibody; Fas ligand receptor; Fas-R; death domain region; septic shock;
KM tumour necrosis factor; HIV-infection; oligodendrocyte death;
KM apoptosis/programmed cell death; p55-R; graft rejection; acute hepatitis;
KM autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;
KM TNF; therapy; ss.
XX
XX Homo sapiens.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 291..1731
FT FT /*tag= a
FT FT /product= MORT-1 binding protein (MACH) isoform alpha
XX
XX WO9703998-A1.
XX
XX 06-FEB-1997.
XX
XX 14-JUN-1996; 96WO-US10521.
XX

```

```

PR 16-APR-1996; 96IL-0117932.
PR 16-JUL-1995; 95IL-0114615.
PR 17-AUG-1995; 95IL-0114986.
PR 14-SEP-1995; 95IL-0115319.
PR 27-SEP-1995; 95IL-0116588.
XX
PA (WEIN/) WEINMURZEL H.
PA (YEDA ) YEDA RES & DEV CO LTD.
XX
PI Boldin M, Goltsev YV, Goncharov T, Wallach D;
XX
DR WPI: 1997-132570/12.
DR P-PSDB: AAM11891.
XX
PT New DNA encoding MACH protein that interacts with MORT-1 protein -
PT to mediate intracellular effects of Fas or TNF receptors, partic.
PT for regulating apoptosis in tumours, virus-infected cells etc.
XX
XX Example 3; Page 116-117; 163pp; English.
XX
XX AAT61396, and AAT61405-T61411 represent coding sequences for different
XX isoforms of MACH. MACH is a binding protein for the mediator of receptor
XX toxicity (MORT-1) protein. MORT-1 binds to the Fas ligand receptor
XX (FAS-R) death domain region, and triggers part of the cell death
XX signalling cascade in mammalian cells. Vectors containing these
XX sequences, the encoded proteins, and antibodies (Ab) against them are
XX used to modulate the effect of Fas-R ligand or TNF on cells that carry
XX Fas-R or p55-R. This is specifically for treating tumours, HIV-infected
XX cells or other diseased cells, by control of apoptosis/programmed cell
XX death. The encoded proteins are mediators of the cell death pathway
XX initiated by TNF and Fas-R binding, i.e. it mimics or enhances the
XX effect of MORT-1 where increased cytotoxicity is required. To inhibit the
XX effect of MORT-1, e.g. in cases of septic shock, graft rejection and
XX acute hepatitis, sequences encoding antisense molecules or ribozymes, or
XX Ab against the protein, are used. Compounds that inhibit MACH are
XX potentially useful for controlling MACH activity e.g. in cases of
XX autoimmune disease, oligodendrocyte death in multiple sclerosis or
XX AIDS-inhibited T-cell suicide. The MACH proteins can also be used to
XX isolate and characterise other proteins and receptors involved in
XX signalling and for Ab production. The Ab can be used to purify the new
XX proteins and for diagnosis of conditions involving abnormal function of
XX Fas-R mediated cellular effects.
XX
SQ Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other:
XX
Query Match      84.8%; Score 17.8; DB 18; Length 2887;
Best Local Similarity 90.5%; Pred. No. 36;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TAGGGGATTGGAGATTGCGA 21
    ||||||| ||||||| |||||
Db 221 TAGGGGACTCGAGACTGCGA 241
XX
RESULT 7
AAD03915
ID AAD03915 standard; DNA; 2887 BP.
XX
AC AAD03915;
XX
DT 02-JUL-2001 (first entry)
XX
DE Human caspase-8 DNA.
XX
XX Human; caspase-8; MACH; FLICE; Mch3;
KM cysteine protease; apoptosis; caspase expression cassette; metastasis;
KM tumour; cathepsin B; urokinase; proliferation; gene therapy; ds.
XX
XX Homo sapiens.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 292..1731
FT FT /*tag= a
FT FT

```

PT Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
 PT prognosing cancer, comprises detecting a modification of genomic DNA
 PT comprising the CASP8 gene that results in inactivation of the gene
 PS Example 3; Page 61; 107pp; English.

XX Investigation of silencing of caspase-8 (CASP8) gene by methylation was
 CC analyzed using methylation-sensitive PCR analysis. Amplification of the
 CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
 CC mixtures containing bisulfite treated DNA. Primers AA51818-23 were
 CC designed to produce a 330 bp fragment in the upstream region of Casp8
 CC gene extending from nucleotides +221 to +541. Wild type primers were
 CC used to amplify the corresponding region of untreated genomic DNA.
 CC Controls without DNA were also performed. CASP8, a cysteine protease, is
 CC part of the death inducing signaling complex (DISC) associated with the
 CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
 CC tumour suppressor gene. The CASP8 promoter region sequences, in
 CC particular Region 1 and Region 2, are crucial to the design and execution
 CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
 CC Methylation PCR can be used to examine even minute amounts of patient
 CC material to demonstrate whether the CASP8 gene expresses an mRNA and
 CC protein product. The CASP8 gene has been localized to human chromosome
 CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
 CC treated by administering a vector that expresses a gene encoding
 CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
 CC tumour in which a myc gene is amplified, such as a neuroblastoma.
 CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
 CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
 CC uterine cervical carcinoma can be diagnosed with the new method. A kit
 CC for screening for a compound that induces death-receptor-mediated
 CC apoptosis in cells containing an inactivated CASP8 gene is also
 CC provided.

XX SQ Sequence 21 BP; 5 A; 4 C; 9 G; 3 T; 0 other;

Query Match 84.8%; Score 17.8; DB 21; Length 21;
 Best Local Similarity 90.5%; Pred. No. 30;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTCGAGATTGCGA 21
 ||||||| ||||||| |||||
 DB 1 TAGGGGACTCGAGACTGCGA 21

RESULT 4

ABN58902
 ID ABN58902 standard; DNA; 60 BP.

AC ABN58902;

DT 15-JUL-2002 (first entry)

DE Human spliced transcript detection oligonucleotide SEQ ID NO:31650.

KW Human; mouse; rat; splice transcript; detection; RNA transcript;

KW splice variant; transcriptome; oligonucleotide library; ss.

OS Homo sapiens.

WO200210449-A2.

PD 07-FEB-2002.

PF 20-JUL-2001; 2001MO-IB01903.

PR 28-JUL-2000; 2000US-221607P.

PR 02-MAY-2001; 2001US-287724P.

PA (COMP-) COMPUGEN INC.

PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

XX WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes.

XX Example 1; SEQ ID 31650; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterizing the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialised mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition, to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 60 BP; 14 A; 8 C; 28 G; 10 T; 0 other;

Query Match 84.8%; Score 17.8; DB 24; Length 60;
 Best Local Similarity 90.5%; Pred. No. 31;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TAGGGGATTCGAGATTGCGA 21
 ||||||| ||||||| |||||
 DB 19 TAGGGGACTCGAGACTGCGA 39

RESULT 5

AAA51793
 ID AAA51793 standard; DNA; 753 BP.

AC AAA51793;

DT 31-OCT-2000 (first entry)

DE CASP8 promoter Region 2.

KW CASP8; caspase-8; promoter; Region 2; inactivation; methylation;

KW Cysteine protease; tumour suppressor; chromosome 2q33-34; neuroblastoma;

KW cancer; death receptor; apoptosis; cytostatic; gene therapy; ss.

OS Homo sapiens.

WO200039347-A1.

PD 06-JUL-2000.

PF 30-DEC-1999; 99MO-US31280.

PR 31-DEC-1998; 98US-0114308.

PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Kidd VT, Lantti JM, Teltz T;

XX WPI; 2000-452423/39.


```
PT comprising the CASP8 gene that results in inactivation of the gene
XX
XX Example 3; Page 61; 107pp; English.
PS
XX
CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was
CC analyzed using methylation-sensitive PCR analysis. Amplification of the
CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
CC mixtures containing bisulfite treated DNA. Primers AA51818-23 were
CC designed to produce a 320 bp fragment in the upstream region of Casp8
CC gene extending from nucleotides +221 to +541. Wild type primers were
CC used to amplify the corresponding region of untreated genomic DNA.
CC Controls without DNA were also performed. CASP8, a cysteine protease, is
CC part of the death inducing signaling complex (DISC) associated with the
CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
CC tumour suppressor gene. The CASP8 promoter region sequences, in
CC particular Region 1 and Region 2, are crucial to the design and execution
CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
CC Methylation PCR can be used to examine even minute amounts of patient
CC material to demonstrate whether the CASP8 gene expresses an mRNA and
CC protein product. The CASP8 gene has been localized to human chromosome
CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
CC treated by administering a vector that expresses a gene encoding
CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
CC tumour in which a myc gene is amplified, such as a neuroblastoma.
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
CC uterine cervical carcinoma can be diagnosed with the new method. A kit
CC for screening for a compound that induces death-receptor-mediated
CC apoptosis in cells containing an inactivated CASP8 gene is also
CC provided.
XX
SQ Sequence 21 BP; 5 A; 2 C; 9 G; 5 T; 0 other;
XX
Query Match 100.0%; Score 21; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TAGGGGATTGCGAGATTGCGA 21
Db 1 TAGGGGATTGCGAGATTGCGA 21
RESULT 2
AA51820
ID AA51820 standard; DNA; 21 BP.
XX
AC AA51820;
XX
DT 31-OCT-2000 (first entry)
XX
DE Sense primer for CASP8 unmethylated-specific bisulfite treated DNA.
XX
XX CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
XX tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
XX death receptor; apoptosis; cytostatic; gene therapy; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200039347-A1.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31280.
XX
PR 31-DEC-1998; 98US-0114308.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Kidd VJ, Lahti JM, Teitz T;
XX
DR WPI; 2000-452423/39.
XX
PT Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
```

```
PT prognosing cancer, comprises detecting a modification of genomic DNA
XX comprising the CASP8 gene that results in inactivation of the gene
XX
XX Example 3; Page 61; 107pp; English.
PS
XX
CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was
CC analyzed using methylation-sensitive PCR analysis. Amplification of the
CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
CC mixtures containing bisulfite treated DNA. Primers AA51818-23 were
CC designed to produce a 320 bp fragment in the upstream region of Casp8
CC gene extending from nucleotides +221 to +541. Wild type primers were
CC used to amplify the corresponding region of untreated genomic DNA.
CC Controls without DNA were also performed. CASP8, a cysteine protease, is
CC part of the death inducing signaling complex (DISC) associated with the
CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
CC tumour suppressor gene. The CASP8 promoter region sequences, in
CC particular Region 1 and Region 2, are crucial to the design and execution
CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
CC Methylation PCR can be used to examine even minute amounts of patient
CC material to demonstrate whether the CASP8 gene expresses an mRNA and
CC protein product. The CASP8 gene has been localized to human chromosome
CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
CC treated by administering a vector that expresses a gene encoding
CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
CC tumour in which a myc gene is amplified, such as a neuroblastoma.
CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
CC uterine cervical carcinoma can be diagnosed with the new method. A kit
CC for screening for a compound that induces death-receptor-mediated
CC apoptosis in cells containing an inactivated CASP8 gene is also
CC provided.
XX
SQ Sequence 21 BP; 5 A; 0 C; 9 G; 7 T; 0 other;
XX
Query Match 84.8%; Score 17.8; DB 21; Length 21;
Best Local Similarity 90.5%; Pred. No. 30;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAGGGGATTGCGAGATTGCGA 21
Db 1 TAGGGGATTGCGAGATTGCGA 21
RESULT 3
AA51822
ID AA51822 standard; DNA; 21 BP.
XX
AC AA51822;
XX
DT 31-OCT-2000 (first entry)
XX
DE Sense primer for untreated CASP8 wild type DNA.
XX
XX CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
XX tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
XX death receptor; apoptosis; cytostatic; gene therapy; primer; ss.
XX
OS Homo sapiens.
XX
PN WO200039347-A1.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31280.
XX
PR 31-DEC-1998; 98US-0114308.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Kidd VJ, Lahti JM, Teitz T;
XX
DR WPI; 2000-452423/39.
XX
XX
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 20.953 seconds
(without alignments)
2257.229 Million cell updates/sec

Title: US-09-477-082-29
Perfect score: 21
Sequence: 1 tagggattcggagattgcga 21

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	AAA51818	Sense primer for C
2	17.8	84.8	21	AAA51820	Sense primer for C
3	17.8	84.8	21	AAA51822	Sense primer for u
4	17.8	84.8	60	ABN8902	Human spliced tran
5	17.8	84.8	753	AAA51793	CASP8 promoter Reg
6	17.8	84.8	2887	AA161405	MACH isoform alpha
7	17.8	84.8	2887	AAD03915	Human caspase-8 DN
8	17.8	84.8	2887	ABK84723	Human cDNA differe
9	17.8	84.8	43938	AAK77216	Human immune/haema

Result No.	Score	Query Match	Length	DB ID	Description
10	17.8	84.8	45017	AAK77217	Human immune/haema
11	17.4	82.9	8951	ABL32794	Human immune syste
12	16.8	80.0	540	ABO22580	Oligonucleotide fo
13	16.8	80.0	540	ABO22581	Oligonucleotide fo
14	16.8	80.0	811	ABO31406	Oligonucleotide fo
15	16.8	80.0	811	ABO31407	Oligonucleotide fo
16	16.8	80.0	5518	ABL54318	Chemically treated
17	16.8	80.0	5518	ABN80019	Human chemically m
18	16.8	80.0	5518	ABL32217	Human immune syste
19	16.4	78.1	1474	ABO33302	Oligonucleotide fo
20	16.4	78.1	1474	ABO33303	Oligonucleotide fo
21	16.2	77.1	40	AAK78888	Human plasmidogen
22	16.2	77.1	358	ABN16542	Human ORF polynuc
23	16.2	77.1	810	ABV10247	Human prostate exp
24	16.2	77.1	1380	AA171344	A. nidulans NADP-d
25	16.2	77.1	1433	AA171343	Aspergillus nidula
26	16.2	77.1	2137	ABL14993	Drosophila melanog
27	16.2	77.1	4809	ABL14992	Drosophila melanog
28	16.2	77.1	5324	AAK78890	Human tissue facto
29	16.2	77.1	7340	AAD28378	Human chemically t
30	16.2	76.2	7517	AAK89061	Human digestive sy
31	15.8	75.2	629	ABN63732	Human cancer relat
32	15.8	75.2	693	ABO13460	Oligonucleotide fo
33	15.8	75.2	693	ABO13461	Oligonucleotide fo
34	15.8	75.2	1642	AAK039010	Arabidopsis thalia
35	15.8	75.2	3063	ABL04176	Drosophila melanog
36	15.8	75.2	5546	ABK34005	Human DNA for stag
37	15.8	75.2	5546	ABL32391	Human immune syste
38	15.8	75.2	6973	ABL33778	Human immune syste
39	15.8	75.2	13627	ABO65976	Human angiogenesis
40	15.4	73.3	619	ABO47022	Oligonucleotide fo
41	15.4	73.3	619	ABO47023	Oligonucleotide fo
42	15.4	73.3	679	AAK77120	DNA encoding novel
43	15.4	73.3	679	AAK77120	DNA encoding novel
44	15.4	73.3	679	AAK77120	DNA encoding novel
45	15.4	73.3	751	ABG38922	Oligonucleotide fo

ALIGNMENTS

RESULT 1
ID AAA51818 standard; DNA; 21 BP.
AC AAA51818;
XX
XX 31-OCT-2000 (first entry)
XX
XX
XX Sense primer for CASP8 methylated-specific disulfite treated DNA.
DE CASP8; caspase-8; promoter; inactivation; methylation; cysteine protease;
XX tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;
KW death receptor; apoptosis; cytosolic; gene therapy; primer; ss.
XX
XX Homo sapiens.
CS
XX
XX W0200039347-A1.
PN
XX
XX 06-JUL-2000.
PD
XX
XX 30-DEC-1999; 99WO-US31280.
PF
XX
XX 31-DEC-1998; 98US-0114308.
PR
XX
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA
XX
XX Kidd VJ, Lahti JM, Teitz T;
PI
XX
XX WPI: 2000-452423/39.
DR
XX
XX Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
PT prognosing cancer, comprises detecting a modification of genomic DNA

sequence is a dinucleotide (TnCa) repeat. At positions 136394 and 136401 a single subclone disagrees with the sequence generated from the spanning PCR product. The subclone calls a T at both positions in contrast to the C called by the PCR product."

BASE COUNT 49306 a 40338 c 39825 g 48802 t
ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 178271;
Best Local Similarity 94.7%; Pred. No. 3e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGC 19
|||||

DB 80859 TAGGGGATTCGAGATTGC 80841

RESULT 14
LOCUS XL007979

DEFINITION Xenopus laevis protein kinase (cdk2) gene, promoter and partial cds.

ACCESSION U07979.1 GI:473584

VERSION U07979.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

promoter

exon

5'UTR

CDS

intron

BASE COUNT

ORIGIN

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

788 a 559 c 667 g 728 t

Query Match 80.0%; Score 16.8; DB 5; Length 2742;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGGGATTCGAGATTGCGA 21
|||||

DB 2483 AGGGGATTCGAGATTGCGA 2502

RESULT 15

LOCUS AX281276

DEFINITION Sequence 18 from Patent WO0177164.

ACCESSION AX281276

VERSION AX281276.1 GI:16608531

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

1383 a 142 c 1452 g 2541 t

Search completed: July 8, 2003, 04:02:36
Job time: 78.7495 secs

Query Match 80.0%; Score 16.8; DB 6; Length 5518;
Best Local Similarity 90.0%; Pred. No. 5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTCGAGATTGCG 20
|||||

DB 46 TAGGGGATTCGAGATTGCG 65

/organism="synthetic construct"

/db_xref="taxon:32630"

/note="chemically treated genomic DNA (Homo sapiens)"

```

misc-feature 5169..5675
/Note="similar to Homo sapiens EST AU144381
(NID:g1005902)"
misc-feature 5194..5761
/Note="similar to Homo sapiens EST BF529656
(NID:g11617019)"
misc-feature 5237..6019
/Note="similar to Homo sapiens EST AU130508
(NID:g1090862)"
misc-feature 5238..6043
/Note="similar to Mus musculus EST BF166846
(NID:g11047198)"
misc-feature 5326..5653
/Note="similar to Homo sapiens EST BG390376
(NID:g13283824)"
misc-feature 5361..6051
/Note="similar to Homo sapiens EST AU135897
(NID:g10996436)"
misc-feature 5361..5812
/Note="similar to EST BG862638 (NID:g14213176)"
misc-feature 5420..6048
/Note="similar to Homo sapiens EST BE788113
(NID:g10209311)"
misc-feature 5551..6327
/Note="similar to Homo sapiens EST AU130771
(NID:g1091125)"
misc-feature 5615..5984
/Note="similar to Homo sapiens EST AA280659 (NID:g1923454)

Query Match 84.8% Score 17.8; DB 9; Length 181150;
Best Local Similarity 90.5%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGGATTGGAGATTGCGA 21
||||| ||||||| |||||
Db 159132 TAGGGGACTGCGAGACTGCGA 159112

RESULT 12
AX345696 8951 bp DNA linear PAT 01-FEB-2002
LOCUS AX345696
DEFINITION Sequence 767 from Patent WO0200928.
ACCESSION AX345696
VERSION AX345696.1 GI:18493582
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
REFERENCE
1
AUTHORS Diek A., Piepenbrock C. and Berlin K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 767 03-JAN-2002;
Epidemiology AG (DE)
FEATURES
source
1..8951
Location/Qualifiers
/organism="synthetic construct"
/db_xref="taxon:32630"
/Note="chemically created genomic DNA (Homo sapiens)"
BASE COUNT 1965 a 348 c 2538 g 3900 t
ORIGIN

Query Match 82.9% Score 17.4; DB 6; Length 8951;
Best Local Similarity 94.7%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GGGGATTGGAGATTGCGA 21
||||| ||||||| |||||
Db 5581 GGGGTTTCGAGATTGCGA 5599

RESULT 13
AC091785/c 178271 bp DNA linear ROD 07-MAR-2002
LOCUS AC091785

```

```

DEFINITION Genomic sequence for Mus musculus, clone RP23-349P1, complete
sequence.
ACCESSION AC091785
VERSION AC091785.5 GI:19224666
KEYWORDS HTG.
SOURCE
MUS musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 178271)
Mccombie, W.R., Spiegel, L., de la Bastide, M., Preston, R.,
Ferraro, K., Kuit, K., Nascimento, L., Zutavern, T., Ballifa, V.,
Bell, M., Baker, J., Santos, L., Miller, B., Katzenberger, F.,
Muller, S., King, L., Yang, C., Dike, S., Palmer, L., O'Shaughnessy, A.
and Dedhia, N.
Genomic sequence for Mus musculus, clone RP23-349P1, complete
sequence
TITLE
Unpublished
2 (bases 1 to 178271)
Mccombie, W.R.
Direct Submission
Submitted (05-JUN-2001) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 178271)
Mccombie, W.R.
Direct Submission
Submitted (07-MAR-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT
On Mar 7, 2002 this sequence version replaced gi:16603978.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
This submission represents a partial sequence of clone RP23-349P1.
Only the unique (non-overlapping) sequence is represented. The
remaining sequence overlaps clone RP23-469C2 (AC098839). An
overlap of 2kb is included in this submission for reference. Base
1 to base 2045 on RP23-349P1 overlaps the sequence of RP23-469C2
from base 190433 to base 192478.
FEATURES
source
1..178271
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-349P1"
/clone_lib="RP23-23"
28760..28970
/Note="We believe the assembly to be correct. The
sequence is covered by a single high quality subclone
sequenced with dye terminator chemistry. There is partial
coverage by three additional subclones with quality values
below phred30."
44235..44270
/Note="We believe the assembly to be correct. The
sequence is covered by a single high quality subclone
sequenced with dye terminator chemistry. There is partial
coverage by an additional subclone with quality below
phred30."
Complement(125470..125510)
/Note="We believe the assembly to be correct. The
sequence is covered by a single high quality subclone
sequenced with dye terminator chemistry. There is partial
coverage by three additional subclones with quality values
below phred30."
136300..136391
/Note="We believe the assembly to be correct. The
sequence is covered solely by a PCR product which was
generated with a high fidelity polymerase. The sequence
is of high quality."
136305..136428
/Note="We believe the assembly to be correct. The

```



```

hypothetical protein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK64751.1"
/db_xref="GI:14523156"
/translation="MERGDKXQENISRLQCAAOQTADPTMOOKATKGEPEKLRFSLR
KRDVWATGALAAALLGIVLIMRLMDVPLRLRNWYAGFLILIMLVANV
LVEFLIGIPNRVSRPENTKRIFPLVVAIVFAIVSLPNNVYAKGFLILILIG
ALOMTISSTIANIYILARAPRVGDRIRIGAHADVDSYDITLMREGEYIMDH
PSGRITKPPNSTVEPTFVNTSWPLFPYVNEIKFQIATSDLEFVARTKREVEEDY
GIMSOAKVIXHILISNTPVDELEKHPVHFVPSVNTLVAIVPLVPKAGRTK
TRIKEMALRMAAEDRVLFPKSNLR"
complement(9248..10165)
/gene="Sma0179"
complement(9248..10165)
/gene="Sma0179"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; similar to glycine cleavage
system transcriptional activator, with bacterial LysR
regulatory helix-turn-helix protein domain"
/codon_start=1
/transl_table=11
/product="putative transcriptional activator"
/protein_id="AAK64752.1"
/db_xref="GI:14523157"

Query Match      84.8%  Score 17.8; DB 1; Length 10253;
Best Local Similarity 90.5%  Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGGATTGGAGATTGCGA 21
      1 ||||| ||||| ||||| |||||
Db      368 TCGGGGATCGGAGATTGCGA 388

RESULT 10
AC0074016      179216 bp      DNA      linear      HTG 09-AUG-2001
LOCUS      Homo sapiens chromosome 2 clone RP11-155P18, WORKING DRAFT
SEQUENCE      2 unordered pieces.
AC0074016
AC0074016.5      GI:15144340
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 179216)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 179216)
Waterston,R.H.
Direct Submission
Submitted (10-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 9, 2001 this sequence version replaced gi:9958173.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: HNH015P18
----- Summary Statistics -----
Sequencing vector: M13; 48%
Sequencing vector: plasmid; 44%
Chemistry: dye-primer ET; 48% of reads
Chemistry: dye-terminator Big Dye; 44% of reads
Assembly program: Phrap; version 0.990319

hypothetical protein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK64751.1"
/db_xref="GI:14523156"
/translation="MERGDKXQENISRLQCAAOQTADPTMOOKATKGEPEKLRFSLR
KRDVWATGALAAALLGIVLIMRLMDVPLRLRNWYAGFLILIMLVANV
LVEFLIGIPNRVSRPENTKRIFPLVVAIVFAIVSLPNNVYAKGFLILILIG
ALOMTISSTIANIYILARAPRVGDRIRIGAHADVDSYDITLMREGEYIMDH
PSGRITKPPNSTVEPTFVNTSWPLFPYVNEIKFQIATSDLEFVARTKREVEEDY
GIMSOAKVIXHILISNTPVDELEKHPVHFVPSVNTLVAIVPLVPKAGRTK
TRIKEMALRMAAEDRVLFPKSNLR"
complement(9248..10165)
/gene="Sma0179"
complement(9248..10165)
/gene="Sma0179"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; similar to glycine cleavage
system transcriptional activator, with bacterial LysR
regulatory helix-turn-helix protein domain"
/codon_start=1
/transl_table=11
/product="putative transcriptional activator"
/protein_id="AAK64752.1"
/db_xref="GI:14523157"

Query Match      84.8%  Score 17.8; DB 2; Length 179216;
Best Local Similarity 90.5%  Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TAGGGGATTGGAGATTGCGA 21
      1 ||||| ||||| ||||| |||||
Db      153057 TAGGGGACTGGAGACTGCGA 153077

RESULT 11
AC007256/c      181150 bp      DNA      linear      PRI 07-NOV-2001
LOCUS      Homo sapiens BAC clone RP11-575C6 from 2, complete sequence.
REFERENCE      AC007256
AC007256.5      GI:14327821
HTG.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 181150)
Waterston,R.H.
Direct Submission
Submitted (07-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE      99063792
JOURNAL      9847074
PUBMED      2 (bases 1 to 181150)
REFERENCE      Du,F., Laplant,Y., Dobbert,A. and Moore,B.
AUTHORS      The sequence of Homo sapiens BAC clone RP11-575C6
TITLE      Unpublished
JOURNAL      3 (bases 1 to 181150)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
JOURNAL      Submitted (07-APR-1999) Genome Sequencing Center, Washington
AUTHORS      University School of Medicine, 444 Forest Park Parkway, St. Louis,
TITLE      MO 63108, USA
JOURNAL      4 (bases 1 to 181150)
REFERENCE

```

JOURNAL Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES Location/Qualifiers

SOURCE

1. 10253 "Sinorhizobium meliloti"

/organism="Sinorhizobium meliloti"

/strain="1021"

/db_xref="taxon:382"

/plasmid="psyma"

/complement(123..308)

/gene="Sma0162"

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK64743.1"

/translation="MIDREKFEISTESPADAGGLTPMFVHRATTKQKNAREPLDL

VWVPELHTRHNVALS"

343..1572

/gene="piliQ2"

/note="Sma0163"

343..1572

/function="Structural elements; Cell exterior; surface

structures"

/note="similar to hypothetical 44.3 kDa protein YXJ in

Rhizobium sp. NGR234"

/transl_table=11

/codon_start=1

/product="probable piliQ pilus assembly protein"

/protein_id="AAK64744.1"

/db_xref="GI:14523149"

/translation="MOREDELVIGDPRIATVPILTORSEFYILGNEIGSTSVTPDAEK

NPVGIIDIEVTDIKLSTIROSVSSYKVSANRITVIGSADPAVATAAEOA

SRKGDSEVINSIKITSSQOVLNAREVEINRDYKGLGTQISAKYMSGSEFFNS

PRATSNTPAGSLISLIGEGSVDAIDALEDRGMARLAEPNLARSGETSSFLAG

EFPIPISEODGTVSYKKEGVGLDFTPTVLSDLILADIEPVSALDNTASVAVGI

AIPGFSARATSVDLKSGQSLAGLQSENNLITQVPLGGLPILGALFSGKAO

RRETDVITVPHVKRPIDPLKVASPSDRTKRTEAEFFLGNIDEVEVNGRSASR

QARVRAPSSGHLEQ"

1569..1865

/gene="Sma0164"

1569..1865

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK64745.1"

/db_xref="GI:14523150"

/translation="MIRSRSPRGFPGLFASISLLSGCQNHLEVRSETTALSAGDAI

AANSVMQVDPMPRVRKQTSLATPADLEQYRPOQPMNEQNGNEETYPNDTTQ"

1887..2417

/gene="Sma0166"

1887..2417

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK64746.1"

/db_xref="GI:14523151"

/translation="MNTTGASGRSVLRSQGRRPTELVAIVATVLLSCQSTSEVL

SGAEDPSPALSSGSDGSKSDGQKLOFMNGNYGLAEKFRKAVELRODAEALMGL

AACVDRGRPLADRAVYQMLKVAQRPRITNNNGYSOYLAGEKAKAKAKLLERAAAS

EGDETIERNLALIDRS"

2501..3313

/gene="Sma0168"

CDS

2501..3313
/gene="Sma0168"

/function="Macromolecule metabolism; Macromolecule

synthesis, modification; RNA synthesis, modification, DNA

transcription"

/note="glimmer prediction; with ribosomal RNA adenine

dimethylases domain"

/codon_start=1

/transl_table=11

/product="methyltransferase-like protein"

/protein_id="AAK64747.1"

/db_xref="GI:14523152"

/translation="MAMASDVLAASFCKEAREGLDPQYHTARAPPELVMLARRAC

LRKGISLEIGAGTGLATERLEDPRHLAVEPDRRLARLGRGLDKELEVEETP

EKLIVPEKFDLVASATAFHWIDAPALRIHLRAGTVALEWVFGDVPDPFH

RATALFSGHRTSPGGGCTTKPYGLNGARLGEAEAGFTADEPELIDMTLADPPA

VRIVATYSNATPALPADERERLSGLEIAETEFAGVYTRNMTTSVYGRRE"

3374..4159

/gene="Sma0169"

3374..4159

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK64748.1"

/db_xref="GI:14523153"

/translation="MALDRADYDAELARHNRQLVAAADFGADRVLDIGGAGGQTR

EAARAPQGEAIGVDISAMLEEARRSAAEGLRNAMPQAGACFGPTGSEDCIS

RCQVMPFAPAAAFNIGRAMRGARLVWQOSERNWRAIRALQALAIYVAGA

ANFSLGDDPPVATDLISAGFTSIDPADQVEYFPGSYDAAFDLITSLYQDALAS

THEPPDRPLQRLDLEGHMTPGVGFDSRAWITARRAGGG"

complement(4340..5104)

/gene="Sma0171"

complement(4340..5104)

/function="Miscellaneous; Unknown"

/note="glimmer prediction"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAK64749.1"

/db_xref="GI:14523154"

/translation="MNAKILLITPILAAFAVYSIVSLYLSORALLYPGASATPAE

RATWQNASIQTPDGETLHGLYSRGPSPVLPFGNDRVSNVGFQALAAKGIG

LIALSYRGYPGSSGTPNEHGLIDGIAADMLAARGENITVVGOSLSGVVDTRAG

RPAVAVIIVSAVLSVLAQTYVPPFVALILTKDPSDLKLAGVRQPAVYPRPARH

HBPIVFGRSSVSRSRAQADAHLCRCPORSVGCPCG"

5321..6241

/gene="Sma0172"

5321..6241

/function="Miscellaneous; Hypothetical/Partial homology"

/note="glimmer prediction; C-terminal homology with

hypothetical protein"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAK64750.1"

/db_xref="GI:14523155"

/translation="MSNSKDEVERIDMLENLADITIDEXELELSPTSEKIREIYR

KAPHPALPRMDYFRALALQAEILKIDWVYKQVYVFEGRDAAGGVAKRTQ

RLNRPRIYVALPAPSDREKTOYFQRYVHLPAAGEIVLEDSYNRKGVKVGFA

TEEEVGEFPDDVPEPEPMIYRSGVRLYKWFESTDEQOLRLTRIHDPLKQKLSPM

DLOSIRVMEVYTRAKETFAKTRIRERPMIVANDKKRRLNCIOHLKQIYEDVP

HEDTLPERIFENYERKVLPEELVPAKI"

complement(6637..7719)

/gene="Sma0175"

complement(6637..7719)

/gene="Sma0175"

/function="Miscellaneous; Hypothetical/Partial homology"

/note="glimmer prediction; partial homology with

TITLE Direct Submission
JOURNAL Submitted (24-MAY-1996) D. Wallach, The Weizmann Institute, Dept of Membrane Research & Biophysics, Rehovot 76100, ISRAEL
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 2887)
AUTHORS Wallach, D.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1996) D. Wallach, The Weizmann Institute, Dept of Membrane Research & Biophysics, Rehovot 76100, ISRAEL
REFERENCE 4 (bases 1 to 2887)
AUTHORS Manduzato, S., Brasseur, F., Andry, G., Boon, T. and van der Bruggen, P.
TITLE A CASP-8 mutation recognized by cytolytic T lymphocytes on a human head and neck carcinoma
JOURNAL T Exp. Med. 186 (5), 785-793 (1997)
MEDLINE 97419196
PUBMED 9271594
FEATURES
source Location/Qualifiers
1..2887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="thymus"
292..1731
/function="protease"
/note="triggers cell death"
/codon_start=1
/product="MACH-alpha-1"
/protein_id="CAA6653.1"
/db_xref="GI:2440071"
/db_xref="SWISS-PROT:Q14790"
/translation="MDFSRNLYDIGEQLDSDSLAKFLSLDYIPQKOEPIKDALML
FORLQKMLERSLNLSFLKELFRINRDLITYLNTKREMERLQPGRAQISAYR
VLYXQISEVSRSELRSFKLQELSKKLDLDDNLDIFLEMKRVYLGEGKLDIL
KVCQAIKMSLIKIIINDYEESEKRSSELSGDPDFSGEELCGVMTSDSPRODSE
SOTLKYQCMKSPKPGCYCLIIINHNFAKAREVPCGLIIRNGKHITDAGATTTFF
LHFETKPHDCTVEQITELIKTYLQMDNSMDCITCGILSHDCKIITIGTQGPATY
ELTSGFTGLKCPSLAGKPRVFTIQACQGDNYQGIPEVTSDEQGYLMDLSSPTRY
IPDEADFLGMATVNNVCYSRNPAQGIWYIOSLCOSLRERCRPGDIIITILEVNEY
SNKDRKMKMGKMPQPTFLTKKLYPSPD"
CDS
BASE COUNT 853 a 595 c 625 g 814 t
ORIGIN
Query Match 84.8%; Score 17.8; DB 9; Length 2887;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAGGGGATTCGGAGATTGCGA 21
||||| |||||| |||||
Db 221 TAGGGACTCGAGACTGCGA 241
RESULT 8
AB038980S2 6502 bp DNA linear PRI 23-FEB-2001
LOCUS Homo sapiens gene for caspase-8, exon 3, exon 4.
DEFINITION Homo sapiens
ACCESSION AB038981.1 GI:12862688
VERSION 2 of 6
KEYWORDS Homo sapiens DNA.
SEGMENT Homo sapiens
SOURCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE Hadano, S., Yanagisawa, Y., Skaug, J., Fichter, K., Nasir, J.,
AUTHORS Martindale, D., Koop, B.F., Scherer, S.W., Nicholson, D.W.,
Roulean, G.A., Ikeda, J.-E. and Hayden, M.R.
TITLE Cloning and characterization of three novel genes, ALS2CR1,
ALS2CR2, and ALS2CR3, in the juvenile amyotrophic lateral sclerosis
(ALS2) critical region at chromosome 2q33-q34: candidate genes for
ALS2
JOURNAL Genomics 71 (2), 200-213 (2001)
MEDLINE 21100893

REFERENCE 2 (bases 1 to 6502)
AUTHORS Hadano, S., Ikeda, J. and Hayden, M.R.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2000) Shinji Hadano, Tokai University, The
Institute of Medical Sciences; Bohseidai, Isehara, Kanagawa
259-1193, Japan (E-mail: shinji@nga.med.u-tokai.ac.jp,
Tel:+81-463-91-5095, Fax:+81-463-91-4993)
FEATURES
source Location/Qualifiers
1..6502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q33"
<1..1478
/note="intron is approximately 24.0 kbp long"
/number=2
1479..1743
/product="caspase-8"
/note="CDS is reported in Acc# AB038985
/alternative 5'UTR"
/number=3
1744..3924
/number=3
3925..3970
/product="caspase-8"
/note="CDS is reported in Acc# AB038985
/alternative 5'UTR"
/number=4
3971..>6502
/note="intron is approximately 5.9 kbp long"
/number=4
BASE COUNT 1718 a 1361 c 1504 g 1919 t
ORIGIN
Query Match 84.8%; Score 17.8; DB 9; Length 6502;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TAGGGGATTCGGAGATTGCGA 21
||||| |||||| |||||
Db 1699 TAGGGACTCGAGACTGCGA 1719
RESULT 9
AE007203 10253 bp DNA linear BCT 15-AUG-2001
LOCUS Sinorhizobium meliloti plasmid pSyma section 9 of 121 of the
DEFINITION Complete plasmid sequence.
ACCESSION AE007203 AE006469
VERSION AE007203.1 GI:14523147
KEYWORDS Sinorhizobium meliloti.
SOURCE Sinorhizobium meliloti.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 10253)
AUTHORS Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P.,
Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J.,
Gurjal, M., Hong, A., Huizaf, L., Hyman, R.W., Kahn, D., Kahn, M.L.,
Kallman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R.,
Wells, D.H., Yeh, K.-C., Davis, R.W., Federspiel, N.A. and Long, S.R.
TITLE Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSyma megaplasmid
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
JOURNAL 21396509
MEDLINE 11481432
PUBMED 11481432
REFERENCE 2 (bases 1 to 10253)
AUTHORS Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P.,
Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J.,
Gurjal, M., Hong, A., Huizaf, L., Hyman, R.W., Kahn, D., Kahn, M.L.,
Kallman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R.,
Wells, D.H., Yeh, K.-C., Davis, R.W., Federspiel, N.A. and Long, S.R.
TITLE Direct Submission

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2503)
 Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M. and Shore, G.C.
 The procaspase-8 isoform, procaspase-8L, recruited to the BAP31
 complex at the endoplasmic reticulum
 Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)

JOURNAL 21927603
 MEDLINE 11917123
 PUBMED 11917123

REFERENCE 2 (bases 1 to 2503)
 Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M. and Shore, G.C.
 Direct Submision
 Submitted (24-SEP-2001) Biochemistry, McGill University, 3655
 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada

JOURNAL
 TITLE Location/Qualifiers
 SOURCE 1..2503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2503
 /gene="CASP8"
 115..1731
 /note="caspase precursor; alternatively spliced"
 /codon_start=1
 /product="procaspase-8L"
 /protein_id="AAL87628.1"
 /db_xref="GI:19401519"
 /translation="MEGRRARVIESRRNPFILGAPTPPEAHVELGRUGDELTAV
 PKRGADYILPEKMPFSRNIDIGQDSEDLASLKFSLDITPRKRPKDALM
 LFORQKRMLESNLSFLKELFRIRLDLITYLTREEMERELQTPGKQISAY
 RVMYQISEVSRSELSRFEKLELRIKLDMDLIDFLEMERVILGSKLDI
 LKRVQAINSLKLIINDYEFSEKSSSESGSPDESNCEELQVWTISDSRQDS
 ESQILDVYQKSKRPKRCYCLINNNHNAKAREKPKHSIRDRNGTILDGALTTFE
 ELHEIRPHDCTVEQIYEILKITYQLDHSNMDQFCICILSHGKGIITGDSQEAR
 YELTSOFTGLKCPSLAGKPRVFTQACQGNQYQGIPEVDSEQPILEMDISSPOR
 YIPEDADFLGMATVNNCVSRNPAEGTWIQLSCSLRRCRPRGDDILITLEVNT
 VSNDDKRNKGKMPQPTFLRKLVPPSD"

BASE COUNT 726 a 528 c 593 g 656 t
 ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 2503;
 Best Local Similarity 90.5%; Pred. No. 1.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCGA 21
 ||||| ||||| ||||| |||||
 Db 221 TAGGGACTCGAGACTGCGA 241

RESULT 5
 AR211527 2887 bp DNA linear PAT 20-JUN-2002
 LOCUS AR211527
 DEFINITION Sequence 14 from patent US 6393937.
 ACCESSION AR211527
 VERSION AR211527.1 GI:21514872
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 2887)
 Wallach, D., Boldin, M., Goncharov, T. and Golstey, V.V.
 TITLE Modulators of the function of Fas receptors and other proteins
 JOURNAL Patent: US 6393937-A 14 04-JUN-2002;
 FEATURES
 source Location/Qualifiers
 1..2887
 /organism="unknown"
 BASE COUNT 853 a 595 c 625 g 814 t
 ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 2887;
 Best Local Similarity 90.5%; Pred. No. 1.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCGA 21
 ||||| ||||| ||||| |||||
 Db 221 TAGGGACTCGAGACTGCGA 241

RESULT 6
 AX134419 2887 bp DNA linear PAT 29-MAY-2001
 LOCUS AX134419
 DEFINITION Sequence 15 from Patent WO0129232.
 ACCESSION AX134419
 VERSION AX134419.1 GI:14271065
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2887)
 Cordell, B. and Li, Y.
 TITLE Functional cloning of genes encoding proteins/enzymes involved in
 proteolytic cleavage
 JOURNAL Patent: WO 0129232-A 15 26-APR-2001;
 Scios Inc. (US)
 FEATURES
 source Location/Qualifiers
 1..2887
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 292..1731
 /note="translated protein product"
 /codon_start=1
 /protein_id="CAC39526.1"
 /translation="MDFSNLYDIGQDSEDLASLKFSLDYIPQKRPKDALM
 LFORQKRMLESNLSFLKELFRIRLDLITYLTREEMERELQTPGKQISAY
 RVMYQISEVSRSELSRFEKLELRIKLDMDLIDFLEMERVILGSKLDI
 LKRVQAINSLKLIINDYEFSEKSSSESGSPDESNCEELQVWTISDSRQDS
 ESQILDVYQKSKRPKRCYCLINNNHNAKAREKPKHSIRDRNGTILDGALTTFE
 ELHEIRPHDCTVEQIYEILKITYQLDHSNMDQFCICILSHGKGIITGDSQEAR
 YELTSOFTGLKCPSLAGKPRVFTQACQGNQYQGIPEVDSEQPILEMDISSPOR
 YIPEDADFLGMATVNNCVSRNPAEGTWIQLSCSLRRCRPRGDDILITLEVNT
 VSNDDKRNKGKMPQPTFLRKLVPPSD"

BASE COUNT 853 a 595 c 625 g 814 t
 ORIGIN

Query Match 84.8%; Score 17.8; DB 6; Length 2887;
 Best Local Similarity 90.5%; Pred. No. 1.4e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TAGGGATTGCGAGATTGCGA 21
 ||||| ||||| ||||| |||||
 Db 221 TAGGGACTCGAGACTGCGA 241

RESULT 7
 HSMACHAL 2887 bp mRNA linear PRI 25-SEP-1997
 LOCUS HSMACHAL
 DEFINITION H sapiens mRNA for MACH-alpha-1 protein.
 ACCESSION X88172
 VERSION X88172.1 GI:1403318
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2887)
 Boldin, M.P., Goncharov, T.M., Golstey, V.V. and Wallach, D.
 TITLE Involvement of MACH, a novel MORT1/FADD-interacting protease, in
 Fas/APO-1- and TNF receptor-induced cell death
 JOURNAL Cell 85 (6), 803-815 (1996)
 MEDLINE 96279826
 PUBMED 8681376

REFERENCE 2 (bases 1 to 2887)
 Wallach, D.

complex at the endoplasmic reticulum
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)

JOURNAL
MEDLINE
21927603
PUBMED
11917123
REFERENCE
2 (bases 1 to 777)
AUTHORS
Breckenridge,D.G., Nguyen,M., Kupfig,S., Reth,M. and Shore,G.C.
TITLE
Submitted (24-SEP-2001) Biochemistry, McGill University, 3655
JOURNAL
Promenade Sir William Osler, Montreal, PQ H3G1V6, Canada

FEATURES
Source
1..777
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="4"
1..>777
/gene="CASP8"
266..>777
/gene="CASP8"
/note="caspase precursor; alternatively spliced"
/codon_start=1
/product="procaspase-8"
/protein_id="AA187632.1"
/db_xref="GI:19401530"
/translation="MDFSRNLYDIGEQLDSEDLASLFLSLDYTPQRKQEPKDALML
FORLOEKRMLESNLSFLKELLPRIINRDLITLNTKREMERLOTQPGRAQISAVR
VMLOYISEVSRSELSRFLQSEISKCKLDDMLDIFIEMERKRVIIIGEGKIDL
KRVCAQINKSL"
KRVCAQINKSL"

CDS
gene
CDS

BASE COUNT
218 a 156 c 209 g 194 t

ORIGIN

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 9; Length 777;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 TAGGGGATTCGAGATTGCGA 21
|||||
221 TAGGGGACTCGGAGACTGCGA 241

Db

RESULT 2
AF210257
LOCUS
AF210257
DEFINITION
Homo sapiens caspase 8 gene, 5' flanking region.
ACCESSION
AF210257.1 GI:8272474
VERSION
AF210257.1
KEYWORDS
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 982)
Teitz,T., Wei,T., Valentine,M.B., Vanin,E.F., Grenet,J.,
Valentine,V.A., Behm,F.G., Look,A.T., Lantti,J.M. and Kidd,V.J.
Caspase 8 is deleted or silenced preferentially in childhood
neuroblastomas with amplification of MYCN
JOURNAL
MEDLINE
20264361
PUBMED
10802708
REFERENCE
2 (bases 1 to 982)
Teitz,T., Wei,T., Valentine,M.B., Vanin,E.F., Grenet,J.,
Valentine,V.A., Behm,F.G., Look,A.T., Lantti,J.M. and Kidd,V.J.
Direct Submission
Submitted (01-DEC-1999) Tumor Cell Biology, St. Jude Children's
Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA

FEATURES
Source
1..982
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q33-q34"
/note="5' flanking region of caspase 8 gene"

misc.feature
1..982
/note="5' flanking region of caspase 8 gene"

BASE COUNT
198 a 235 c 241 g 308 t

ORIGIN

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 9; Length 982;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 TAGGGGATTCGAGATTGCGA 21
|||||
536 TAGGGGACTCGGAGACTGCGA 556

Db

RESULT 3
AF380342
LOCUS
AF380342
DEFINITION
Homo sapiens caspase-8L mRNA, complete cds.
ACCESSION
AF380342.1 GI:14211397
VERSION
AF380342.1
KEYWORDS
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1723)
Himeji,D., Horiuchi,T., Tsukamoto,H., Hayashi,K., Watanabe,T. and
Harada,M.
TITLE
Characterization of caspase-8L: a novel isoform of caspase-8 that
behaves as an inhibitor of the caspase cascade
JOURNAL
MEDLINE
22005982
PUBMED
12010809
REFERENCE
2 (bases 1 to 1723)
Himeji,D., Tsukamoto,H. and Horiuchi,T.
AUTHORS
Direct Submission
TITLE
Submitted (10-MAY-2001) Medicine and Biosystemic Science, Kyushu
University Graduate School of Medical Science, 3-1-1 Maidashi,
Higashi-Ku, Fukuoka 812-8582, Japan

FEATURES
Source
1..1723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="peripheral blood lymphocyte"
86..916
/codon_start=1
/product="caspase-8L"
/protein_id="AAK57457.1"
/db_xref="GI:14211398"
/translation="MDFSRNLYDIGEQLDSEDLASLFLSLDYTPQRKQEPKDALML
FORLOEKRMLESNLSFLKELLPRIINRDLITLNTKREMERLOTQPGRAQISAVR
VMLOYISEVSRSELSRFLQSEISKCKLDDMLDIFIEMERKRVIIIGEGKIDL
KRVCAQINKSLKINIDYEFSEKRSSESGSDPDSNGEGLGWTIISDPEDQSE
SQTLDKVYQMKSRKPRGICLTINHNFRKAREKVPKLSHIDRNGTHLDAGTVEPKREK

CDS
caspase-8L

BASE COUNT
538 a 344 c 411 g 430 t

ORIGIN

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 9; Length 1723;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY
1 TAGGGGATTCGAGATTGCGA 21
|||||
15 TAGGGGACTCGGAGACTGCGA 35

Db

RESULT 4
AF422925
LOCUS
AF422925
DEFINITION
Homo sapiens procaspase-8L (CASP8) mRNA, complete cds;
ACCESSION
AF422925.1 GI:19401518
VERSION
AF422925.1
KEYWORDS
Homo sapiens.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 74.5828 Seconds

(without alignments)
8194.363 Million cell updates/sec

Title: US-09-477-082-29

Sequence: 1 tagggagtcgagcattgcga 21

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: 1: gb_da: 2: gb_hcg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sts: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: em_da: 16: em_fun: 17: em_hum: 18: em_in: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pat: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sts: 28: em_un: 29: em_vl: 30: em_hcg_hum: 31: em_hcg_inv: 32: em_hcg_other: 33: em_hcg_mus: 34: em_hcg_pin: 35: em_hcg_rtd: 36: em_hcg_mam: 37: em_hcg_vrt: 38: em_sy: 39: em_hgo_hum: 40: em_hgo_mus: 41: em_hgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.8	84.8	777	9	AF422929 Homo sapi
2	17.8	84.8	982	9	AF210257 Homo sapi
3	17.8	84.8	1723	9	AF380342 Homo sapi
4	17.8	84.8	2503	9	AF422925 Homo sapi
5	17.8	84.8	2887	6	AR211527 Sequence
6	17.8	84.8	2887	6	AX134419 Sequence
7	17.8	84.8	2887	6	AX134419 Sequence
8	17.8	84.8	6502	9	AB038980S2
9	17.8	84.8	10253	1	AE007203
10	17.8	84.8	179216	2	AC074016 Homo sapi
11	17.8	84.8	181150	5	AC007256 Homo sapi
12	17.4	82.9	8651	6	AX345696 Sequence
13	17.4	82.9	178271	10	AC091785
14	16.8	80.0	2742	5	XL007979
15	16.8	80.0	5518	6	AX281276
16	16.8	80.0	5518	6	AX344611 Sequence
17	16.8	80.0	5518	6	AX345119 Sequence
18	16.8	80.0	141382	2	CNS0808K
19	16.8	80.0	148783	2	AC097575
20	16.8	80.0	151944	2	AC115475
21	16.4	78.1	2785	3	AB007406
22	16.4	78.1	38687	3	U23515
23	16.4	78.1	70122	9	AC091938
24	16.4	78.1	116211	8	AP003843
25	16.4	78.1	144561	2	AP004307
26	16.4	78.1	150699	2	AC099245
27	16.4	78.1	179686	2	AC023987
28	16.4	78.1	191052	10	AL683814
29	16.2	77.1	290	9	HS21H7F
30	16.2	77.1	1759	3	AF177464
31	16.2	77.1	7618	1	AE005734
32	16.2	77.1	20303	3	AC006749
33	16.2	77.1	59527	2	AC014163
34	16.2	77.1	63268	2	AC103683
35	16.2	77.1	128631	8	AP003379
36	16.2	77.1	133084	2	AC099208
37	16.2	77.1	171705	2	AC104166
38	16.2	77.1	171556	2	AC074388
39	16.2	77.1	176145	2	AL731550 Homo sapi
40	16.2	77.1	176444	2	AC130018
41	16.2	77.1	184271	3	AC009256
42	16.2	77.1	186184	2	AC107275
43	16.2	77.1	191709	2	AC097846
44	16.2	77.1	193444	8	AF229199
45	16.2	77.1	193603	2	AC093932

ALIGNMENTS

RESULT 1
AF422929 777 bp mRNA linear PRI 03-APR-2002
LOCUS Homo sapiens clone 4 procaspase-8 (CASP8) mRNA, partial cds;
DEFINITION alternatively spliced.

ACCESSION AF422929
VERSION AF422929.1 GI:19401529

KEYWORDS
SOURCE
ORGANISM

Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 (bases 1 to 777)
Breckenridge, D.G., Nguyen, M., Kuppig, S., Reth, M. and Shore, G.C.
The procaspase-8 isoform, procaspase-8L, recruited to the BAP31

Query Match	4.6%	Score 34.6;	DB 9;	Length 9025608;
Best Local Similarity	55.4%;	Pred. No. 93;		
Matches 67;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0

QY	275	CCGAGAGTCTCCTGTGTGTGTTCTCTCTCTGAGCCGATGCTTTGACTTTGCTACTTTTTCAC	334
Db	8877178	CTCTGTGTGTGCGCTGTGTCTGTCTCTGTGCGGCGCGCTGCGCTTCTTCTCTCTTCCG	8877119
QY	335	TCTGAGCAGTCTCCAGTTCCTCTCTGCTACCTTTTGTCTCCAGAGTTTCCCGCGCGCTCG	394
Db	8877118	CGTGCGCCCGTGTGCGCTTCTCTCTCTGACTCTCTCTTCTCACTTCCCGCTTCCCTCTCG	8877058
QY	395	A	395
		A	
Db	8877058	A	8877058

QY 355 CTGTGACCTTTTGTCTCCAGAGTCCGCTCCGCTCGAGATGAGATACAGAGACTCC 414
 Db 359 .Y.CCGN.H.YNR....H.TKSM.DDD.AGCS.M.BVCN.WGM.M...RACVA. 330
 QY 415 CTCTGTGAGACCCGTTTGGAGAGTCGACAGAGACTTTATCATCAGCTTTTTCCTTTT 474
 Db 329 MBSSPAHNSTG.SYHW....G.YCH....TWYSS.AYN.C.YH.D.T.B.SKH... 270
 QY 475 CATTTGGCCCTGGGGGCGGACGGTTAAGACTTTTCTGTCATCTGTGCGAATCAGAA 534
 Db 269 M..CG.S.CGY.BTC.TCC.B.YTCSTSB-.Y.CTBSHT.YKY.HHMS.WS.T 212
 QY 535 TGCCCTGAGGTGCACAGGCCCTTTCCTTCCTTTTCGCGTCTGTAAGGGGCTTTCCCTTTTANG 594
 Db 211 S.SR.Y.NS.YSS...H.CCTYBC.CYE.YW.Y.C.MB.NTC.TRC.TBTH.YTT... 152
 QY 595 TCTTCACCCGACCCCTTTCCTCCCTCCCTCCCTCGCTGT 631
 Db 151 .S.CB.T.C.TTBTBDMTYIM.MBACY.NT.MSY 115

RESULT 15
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SAOTSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBATA, TADAYOSHI

RESULT 14
US-10-123-155-210/c
; Sequence 210, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:


```

1 ORGANISM: Homo sapiens
2 FEATURE:
3 NAME/KEY: misc.feature
4 LOCATION: 280, 324, 379, 383, 395, 397, 399, 407, 413, 427, 431, 432
5 LOCATION: 433, 445, 458, 465, 474, 479, 499, 512, 519, 520, 526, 527
6 LOCATION: 529, 533, 540, 541, 559, 576, 577, 604, 605, 609, 610
7 LOCATION: 617, 623, 624, 628, 639, 659, 660, 666, 667, 672, 679
8 OTHER INFORMATION: n = A,T,C or G
9 FEATURE:
10 NAME/KEY: misc.feature
11 LOCATION: 680, 682, 684, 685, 686, 687, 688, 689, 711, 712, 713, 714
12 LOCATION: 715, 718, 721, 723, 740, 751, 752, 753, 755, 756, 757, 758
13 LOCATION: 759, 760, 761, 772, 775, 779, 783, 784, 787, 788, 789, 790
14 LOCATION: 792, 794, 795, 797, 802, 809, 814, 815, 816, 817, 818
15 OTHER INFORMATION: n = A,T,C or G
16 FEATURE:
17 NAME/KEY: misc.feature
18 LOCATION: 818, 820, 821, 822, 828, 830, 831, 832, 833, 834, 835, 839
19 LOCATION: 840, 841, 842, 843, 845, 846, 847, 849, 851, 855, 857, 862
20 LOCATION: 865, 868, 869, 886, 900, 901
21 OTHER INFORMATION: n = A,T,C or G
22 IS-10-198-846-7264

```

RESULT 8
 US-10-295-403-75/c
 : Sequence 75, Application US/10295403
 : Publication No. US20030101481a1
 : GENERAL INFORMATION:
 : APPLICANT: Heard, Jacqueline
 : APPLICANT: Riechmann, Jose Luis
 : APPLICANT: Adam, Luc
 : APPLICANT: Brown, Pierre
 : APPLICANT: Pineda, Omaira
 : APPLICANT: Reuber, Lynne
 : APPLICANT: Jiang, Cai-Zhong
 : APPLICANT: Keddie, James
 : APPLICANT: Zhang, James
 : APPLICANT: Benito, Maria-Ines
 : APPLICANT: Yu, Guo-Iiang
 : APPLICANT: Fromm, Mike
 : TITLE OF INVENTION: PLANT GENE SEQUENCES
 : FILE REFERENCE: MBI-0003
 : CURRENT APPLICATION NUMBER: US/10/295,403

Db 742 .G..M.T.H.TYBESRD.SH..KB...T.DTHCHT...T.HS..NABK...C.B..CS. 683
 QY 320 TTGCTACTTTTTCAGTGTGAGCAGTCCAGTCTGCTACTCTTTTCTCTCCAGC 379
 Db 682 T..TAACHBHK.....ASHBYAG.YTDB...NS.G.MH.CGC.M..D..TY.AS..DTN 623
 QY 380 TTCCCTGCGCCCTCGAATGAGATACAGAGTCCCTTGTGTGAGCCCGTTGGAGATC 439
 Db 622 ...TB...C.C.T.Y.Y.A..SRS.ABB.TY.MHBN.....KTHGBYHD..DM.WBAC 563
 QY 440 CAGAGACTTTATTCATTCATCTGTGAGATGAGTCCCTGAGTCCAGTCCCTTTC 499
 Db 562 ST.DM...NYS..ABY.B.CY.YYHATH.BRH.TBA.SRSS.BT.RS.GBY.HBGBSD 503
 QY 500 AAGTACTTTATTCATTCATCTGTGAGATGAGTCCCTGAGTCCAGTCCCTTTC 559
 Db 502 CCTGN...H.SY.T.SB.DSBD...YTY..BTHT.SCB..H.R.NCHKH.SCS.SRC. 443
 QY 560 CCCTCTTTCGCGCTCGAAGGGGTTTCCTTTATGTCTTCCAGCCCGTTCCTTC 619
 Db 442 SM.CBT.C..B.S.MTBBC.SSCDR..BMCHCH..TCBT.H..CH.N.HC.BT.S.C.KS 383
 QY 620 C 620
 Db 382 H 382

RESULT 2

US-10-184-634-120/C
 ; Sequence 120, Application US/10184634
 ; Publication No. US20030068684A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C217
 ; CURRENT APPLICATION NUMBER: US/10/184,634
 ; CURRENT FILING DATE: 2002-06-28
 ; Prior Application removed - See file wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 120
 ; LENGTH: 1141
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-10-184-634-120

Query Match 5.4%; Score 40.6; DB 9; Length 1141;
 Best Local Similarity 11.0%; Pred. No. 0.011;
 Matches 53; Conservative 145; Mismatches 281; Indels 2; Gaps 1;

QY 142 CCTGTGCGAGGTGGGCTCTTCAACAGAAACCAATATTTTGTGTGCTGCT 201
 Db 862 SC.AY.S.C.NSBB.RMKSCBHY.S.KTY.CYBBS.....TKCT.S.....BYT 803
 QY 202 CTAGAACAGGGGTGTGGGGGTGGGAGCACTTGATGCTGCCCTTCTGAGACACCTC 261
 Db 802 YTSB.D...S.AT...B...AA..S.CSAS...R.AB.TCYMK.NC..G.BDSTN.NS 743
 QY 262 TGGTGTGCGCGTGGCCAGGTCTCTGTGTGTGTTTCTCTGAGCCGATG--CCTTGACT 319
 Db 742 .G..M.T.H.TYBESRD.SH..KB...T.DTHCHT...T.HS..NABK...C.B..CS. 683

QY 320 TTGCTACTTTTTCAGTGTGAGCAGTCTCCAGTTCCTCTGCTACTTTTGTCTCCAGC 379
 Db 682 T..TAACHBHK.....ASHBYAG.YTDB...NS.G.MH.CGC.M..D..TY.AS..DTN 623
 QY 380 TTCCCTGCGCCCTCGAATGAGATACAGAGTCCCTTGTGTGAGCCCGTTGGAGATC 439
 Db 622 ...TB...C.C.T.Y.Y.A..SRS.ABB.TY.MHBN.....KTHGBYHD..DM.WBAC 563
 QY 440 CAGAGACTTTATTCATTCATCTGTGAGATGAGTCCCTGAGTCCAGTCCCTTTC 499
 Db 562 ST.DM...NYS..ABY.B.CY.YYHATH.BRH.TBA.SRSS.BT.RS.GBY.HBGBSD 503
 QY 500 AAGTACTTTATTCATTCATCTGTGAGATGAGTCCCTGAGTCCAGTCCCTTTC 559
 Db 502 CCTGN...H.SY.T.SB.DSBD...YTY..BTHT.SCB..H.R.NCHKH.SCS.SRC. 443
 QY 560 CCCTCTTTCGCGCTCGAAGGGGTTTCCTTTATGTCTTCCAGCCCGTTCCTTC 619
 Db 442 SM.CBT.C..B.S.MTBBC.SSCDR..BMCHCH..TCBT.H..CH.N.HC.BT.S.C.KS 383
 QY 620 C 620
 Db 382 H 382

RESULT 3

US-10-123-155-144/C
 ; Sequence 144, Application US/10123155
 ; Publication No. US20030068794A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Ellavort, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tumas, Daniel
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C30
 ; CURRENT APPLICATION NUMBER: US/10/123,155
 ; CURRENT FILING DATE: 2002-04-15
 ; Prior Application removed - See Palm or file wrapper
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 144
 ; LENGTH: 520
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 ; US-10-123-155-144

Query Match 5.1%; Score 38.6; DB 9; Length 520;
 Best Local Similarity 10.1%; Pred. No. 0.034;
 Matches 49; Conservative 135; Mismatches 301; Indels 0; Gaps 0;

QY 257 ACCTGTGCTGCTGCGCCAGGTCTCTGTGTGTTTCTCTGAGCGCATGCTTG 316
 Db 485 RVHHY...SCH.Y..TCB...CCWH..SY.RR..B..CCYT.BRYTCM...YY... 426
 QY 317 ACTTGTCTACTTTTACATCTGTGAGATGAGTCCAGTCCCTGCTACTTTTGTCTCTCA 376
 Db 425 ...TTH.DV...CCTTG...CWB.M.CY.YWGYCM..R.HCHN.SBTTT..N..CWB 366
 QY 377 AGCTTCCCTGCGCCCTCGAATGAGATACAGAGTCCCTTGTGTGAGACCCGTTGAGA 436


```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Mb
: COMPUTER: IBM compatible PC
: OPERATING SYSTEM: Windows 95
: SOFTWARE: Word 97 (DOS text format)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/018,584A
: FILING DATE: 04-Feb-1998
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Grady J. Frenchick
: REGISTRATION NUMBER: 29,018
: REFERENCE/DOCKET NUMBER: 16026,9180
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (608) 257-3501
: TELEFAX: (608) 257-2275
: INFORMATION FOR SRO ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 444 bp
: TYPE: Nucleic Acid
: STRANDEDNESS: Double
: TOPOLOGY: Circular
: MOLECULE TYPE: Genomic DNA
: HYPOTHETICAL: no
: IMMEDIATE SOURCE:
: LIBRARY: plasmid, pGem3zf(+)
: CLONE: G210
: POSITION IN GENOME:
: CHROMOSOME/SEGMENT: 8p
: US-09-018-584A-19

```

```

Query Match          4.2%; Score 31.4; DB 4; Length 444;
Best Local Similarity 64.4%; Pred. No. 2.2;
Matches 47; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 584 TTGCTTTATGTTGTCACCCGACCCCTTCCCTCCCTGCTGTTTGTGCGCA 643
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 335 TTCTTTCTCTCTCCCTCCCTCCCTCCCTCCCTCTCTCTCTCTCTCTCTCT 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 644 AAAACCAATGCT 656
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 395 AAAACCAATGCT 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: July 8, 2003, 05:55:08
 Job time : 198.615 secs

REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 INFORMATION FOR SEQ ID NO: 323:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1166 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 US-09-072-556-323

Query Match 4.4%; Score 32.8; DB 4; Length 1166;
 Best Local Similarity 27.1%; Pred. No. 1.4;
 Matches 102; Conservative 73; Mismatches 197; Indels 4; Gaps 1;

QY 270 COTGGCCAGGCTCTGCTGTGCTGTCTGTGAGCCGATGCTTGTGCTTGT 329
 ||| : : : : : ||| : : : : :
 DB 728 CCMNMCMTGCTWNAKCANONCTTCTCTCMYMTMACMCNNCCKSGACCT 787
 ||| : : : : : ||| : : : : :
 QY 330 TTCACTGTAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 389
 ||| : : : : : ||| : : : : :
 DB 788 CTCACCTKCKCKMCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 847
 ||| : : : : : ||| : : : : :
 QY 390 CCTGCAATGCAATACAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 449
 ||| : : : : : ||| : : : : :
 DB 848 MMYCYTAAKCAKCMKCTCCCKANMKACAKCTCTCCCKANMKACCKCCCTCT 907
 ||| : : : : : ||| : : : : :
 QY 450 TATCAATCACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 505
 ||| : : : : : ||| : : : : :
 DB 908 TCCWCTCTCCTCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 967
 ||| : : : : : ||| : : : : :
 QY 506 TTTATTTCTGATTTCTGATGATGATGATGATGATGATGATGATGATGATG 565
 ||| : : : : : ||| : : : : :
 DB 968 TCTCTNYCTCWCAGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027
 ||| : : : : : ||| : : : : :
 QY 566 TTCGGTCTGAGGAGGTTCTTTTATGCTTTCACCCCACTTTCCCTCCCTG 625
 ||| : : : : : ||| : : : : :
 DB 1028 CTCTCTCTCMKMKCTCCCKCATCTMCKSTCTCTCTCTCTCTCTCTCTCT 1087
 ||| : : : : : ||| : : : : :
 QY 626 CTCTGTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 641
 ||| : : : : : ||| : : : : :
 DB 1088 MYGCTTCT 1103
 ||| : : : : : ||| : : : : :
 QY
 DB

RESULT 11
 US-08-921-195-1/C
 ; Sequence 1, Application US/08921195
 ; Patent No. 6147052
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAU, RAYMOND M.W.
 ; TITLE OF INVENTION: ISOLATION AND USE OF ERYTHROID
 ; TITLE OF INVENTION: DIFFERENTIATION AND DENUCLEATION FACTORS
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LYON & LYON
 ; STREET: 633 WEST SIXTH ST., SUITE 4700
 ; CITY: LOS ANGELES
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 90071
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/921.195
 ; FILING DATE: 29-AUG-1997
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:

NAME: SCHNEIDER, CAROL A.
 REGISTRATION NUMBER: 34,923
 REFERENCE/DOCKET NUMBER: 227/183
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 213-489-1600
 TELEFAX: 213-955-0440
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2721 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-921-195-1

Query Match 4.3%; Score 32.4; DB 3; Length 2721;
 Best Local Similarity 47.9%; Pred. No. 2.9;
 Matches 93; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 282 CTCTGTGTGCTTCTCTGTGAGCCGATGCTTGTGCTTGTGCTTGTGCTG 341
 ||| : : : : : ||| : : : : :
 DB 1534 CTCCTGGGTGATTTACCTACAGCTGAGCTTCTGCTGTGAGCCACTGAC 1475
 ||| : : : : : ||| : : : : :
 QY 342 AGCTCCAGTCTCTGCTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTG 401
 ||| : : : : : ||| : : : : :
 DB 1474 TATGCTCCAGTCCAGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1415
 ||| : : : : : ||| : : : : :
 QY 402 ATACAGGACTCCCTCTGTGAGCCGTTGAGAGTCCAGAGACTTATCATC 461
 ||| : : : : : ||| : : : : :
 DB 1414 CCACCTGAGCCGCGGACGCTCAGTCCCTGCGAGTCCAGAGACTTGTG 1355
 ||| : : : : : ||| : : : : :
 QY 462 TTTTCTTTCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 475
 ||| : : : : : ||| : : : : :
 DB 1354 GTATCCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1341
 ||| : : : : : ||| : : : : :
 QY
 DB

RESULT 12
 US-08-785-420-1/C
 ; Sequence 1, Application US/08785420
 ; Patent No. 6001976
 ; GENERAL INFORMATION:
 ; APPLICANT: MacLennan, David H
 ; APPLICANT: O'Brien, Peter J.
 ; TITLE OF INVENTION: DIAGNOSIS FOR PORCINE MALIGNANT
 ; TITLE OF INVENTION: HYPERHERMIA
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Bell, Seltzer, Park & Gibson
 ; STREET: P.O. Drawer 34009
 ; CITY: Charlotte,
 ; STATE: NC 6001976th Carolina 28234
 ; COUNTRY: U.S.A.
 ; ZIP: 28234
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/785.420
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/249,388
 ; FILING DATE:
 ; APPLICATION NUMBER: US 08/030,159
 ; FILING DATE: 15-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Layton, Jr., Samuel G.
 ; REGISTRATION NUMBER: 22807
 ; REFERENCE/DOCKET NUMBER: 3477-73
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 704-377-1561

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,613
FILING DATE: 30-NOV-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Bucca Ph.D., Daniel
REGISTRATION NUMBER: 42,368
REFERENCE/DOCKET NUMBER: 5019-061
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-756-8000
TELEFAX: 202-756-8087
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2478 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-194-613-1

Query Match 4.4%; Score 33.4; DB 4; Length 2478;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 25 AAGAACTTCTTCTGAGAGCTTTCCACCCCTTCCCTGCTGAGCAGGTAGG 84
DB 2177 AAGGACTTGATGATGAGAGAGCTTCCCTCCGACTTCAGCAGGAGAGAGATGTGCG 2118
QY 85 CAGGTAGGAGACTCGGAGACTGCGATGTGCTCCAGAGAAAGGTGAGCGGTGAGTGCCT 144
DB 2117 GCGGAGGAGAGAAAGCGCGCGCGGAGAGTGTGACGTGCGGTAGACACCTGTGAGACG 2058
QY 145 GTTGCCAGGTGGCTCTTCAACAGAGAAACCAATATTTTG 187
DB 2057 GGTGGCCAGGTGCTGTGCGGAGGAGAGATCTCTCATTTCTTG 2015

RESULT 9
US-08-485-355B-51/c
Sequence 51, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehn Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU P14081/92

FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 2479 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 283..2307
SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-08-485-355B-51

Query Match 4.4%; Score 33.4; DB 4; Length 2479;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 25 AAGAACTTCTTCTGAGAGCTTTCCACCCCTTCCCTGCTGAGCAGGTAGG 84
DB 2178 AAGGACTTGATGATGAGAGAGCTTCCCTCCGACTTCAGCAGGAGAGACATGTGCG 2119
QY 85 CAGGTAGGAGACTGAGAGACTGCGATGTGCTCCAGAGAAAGGTGAGCGGTGAGTGCCT 144
DB 2118 GCGGAGGAGAGAAAGCGCGCGGAGAGTGTGACGTGCGGTAGACACCTGTGAGACG 2059
QY 145 GTTGCCAGGTGGCTCTTCAACAGAGAAACCAATATTTTG 187
DB 2058 GGTGGCCAGGTGCTGTGCGGAGGAGAGATCTCTCATTTCTTG 2016

RESULT 10
US-09-072-596-323
Sequence 323, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.

ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 2478 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 283..753
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-08-485-355B-47
Query Match 4.4%; Score 33.4; DB 4; Length 2478;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 25 AAGAACTTCTTCCTGGAGCCTTCCACCCCTTCCTGCTGAGCAGCGTAGTGG 84
DB 2177 AACGACTTGTATGATGAGGAGCGTTGCTCGACTTCAGCAGGAGGAGCATGTTGCG 2118
QY 85 CAGGTAGGAGCTCGAGACTGCGATGTCAGAGAAAGGTGAGCGGCTGAGTGCCT 144
DB 2117 GCGGAAGGAGCAAGGCGGCGCGGAAGTTGTCAGTGCGGCGGTAGACACTGTGAGACG 2058
QY 145 GTTGCCAGAGGTGGCTCTTCAACAGGAAACCAATATTTTGG 187
DB 2057 GGTGGCCAGGTGCTGCGGAGGAGGATCTCCTCATTTCTTG 2015
RESULT 7
US-08-485-355B-49/c
Sequence 49, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco

STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 2478 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 366..2306
SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-08-485-355B-49
Query Match 4.4%; Score 33.4; DB 4; Length 2478;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 25 AAGAACTTCTTCCTGGAGCCTTCCACCCCTTCCTGCTGAGCAGCGTAGTGG 84
DB 2177 AACGACTTGTATGATGAGGAGCGTTGCTCGACTTCAGCAGGAGGAGCATGTTGCG 2118
QY 85 CAGGTAGGAGCTCGAGACTGCGATGTCAGAGAAAGGTGAGCGGCTGAGTGCCT 144
DB 2117 GCGGAAGGAGCAAGGCGGCGCGCGGAAGTTGTCAGTGCGGCGGTAGACACTGTGAGACG 2058
QY 145 GTTGCCAGAGGTGGCTCTTCAACAGGAAACCAATATTTTGG 187
DB 2057 GGTGGCCAGGTGCTGCGGAGGAGGATCTCCTCATTTCTTG 2015
RESULT 8
US-09-194-613-1/c
Sequence 1, Application US/09194613
Patent No. 6251654
GENERAL INFORMATION:
APPLICANT: GORDON, Karl H.
APPLICANT: HANZLIK, Terry N.
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 600 13th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3096

```
OY 401 GATACAGGACTCCCTTCTGTGACCCCTTTGGAGATCCAGAACACTTATATCAATCCAC 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1191 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1250
OY 461 TTTTTCCTTTTCATTTGGCCCGGGGCGGAGCGTAGTACTTATCTGTGATCC 520
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1251 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1310
OY 521 TGTGCAACAGCAAGTCCCTGAGTGCACAGCCCTTCCCTCTTTCGCGTGAAG 580
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1311 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1370
OY 581 GGTTCCTTTATGCTCTCCACCCCACTTCCCTCCCTGCTGTTTGTGTC 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1371 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1430
OY 641 CCAAAAACAGTCTCTCAACGTTT 667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1431 YYYGTACCAATTTCTTATCTCTT 1457
```

RESULT 4

```
US-09-007-005-17/c
; Sequence 17, Application US/09007005B
; Patent No. 628558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihne
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

Query Match Best Local Similarity 4.9%; Score 37.2; DB 4; Length 289;

Matches 36; Conservative 97; Mismatches 149; Indels 0; Gaps 0;

```
OY 180 TATTTTGTCTGACTGCTCTGAAACAGGCGCTGGGGGTGGGAGCACTTGA 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 TTTTTCCTTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
OY 240 TGTGCGCTCTGAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 CTYGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 165
OY 300 CTGAGCCGATGCTTGTGACTTGTGACTTGTGACTTGTGACTTGTGACTTGTG 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 SYNNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSY 105
OY 360 TACCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 SYNNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSY 45
```

```
OY 420 GTGAGCCGTTTGGAGATCCAGAGACTTATATCAATCCACT 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 44 AYAATYTYGYTAAVATYAAVATYAAVATYAAVATYTYGYTCYCY 3
```

RESULT 5

```
US-09-244-796-17/c
; Sequence 17, Application US/09244796
; Patent No. 628134
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihne
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17
```

Query Match Best Local Similarity 4.9%; Score 37.2; DB 4; Length 289;

Matches 36; Conservative 97; Mismatches 149; Indels 0; Gaps 0;

```
OY 180 TATTTTGTCTGACTGCTCTGAAACAGGCGCTGGGGGTGGGAGCACTTGA 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 TTTTTCCTTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
OY 240 TGTGCGCTCTGAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 CTYGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 165
OY 300 CTGAGCCGATGCTTGTGACTTGTGACTTGTGACTTGTGACTTGTGACTTGTG 359
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 SYNNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSY 105
OY 360 TACCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 SYNNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSYNTSY 45
```

RESULT 6

```
US-08-485-355B-47/c
; Sequence 47, Application US/08485355B
; Patent No. 617075
; GENERAL INFORMATION:
; APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; Protecting Plants
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 ; Search time 180.615 Seconds
(without alignments)
1278.561 Million cell updates/sec

Title: US-09-477-082-2
Perfect score: 753
Sequence: 1 aatagaccgcgcatgtgaaa.....taactcgttttttaacctt 753

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapept 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2.6/prodata/1/lna/5A_COMB.seq:*
2: /cgn2.6/prodata/1/lna/5B_COMB.seq:*
3: /cgn2.6/prodata/1/lna/5C_COMB.seq:*
4: /cgn2.6/prodata/1/lna/5D_COMB.seq:*
5: /cgn2.6/prodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2.6/prodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	123.6	16.4	2887	US-08-883-502-14	Sequence 14, Appl
2	123.6	16.4	2887	PCT-US96-10521-14	Sequence 14, Appl
3	83.8	11.1	7218	US-08-332-463-14	Sequence 14, Appl
4	37.2	4.9	289	US-09-007-005-17	Sequence 17, Appl
5	37.2	4.9	289	US-09-244-796-17	Sequence 17, Appl
6	33.4	4.4	2478	US-08-485-355-47	Sequence 47, Appl
7	33.4	4.4	2478	US-08-485-355-49	Sequence 49, Appl
8	33.4	4.4	2478	US-09-194-613-1	Sequence 1, Appl
9	33.4	4.4	2478	US-08-485-355B-51	Sequence 51, Appl
10	32.8	4.4	1166	US-09-072-596-323	Sequence 323, App
11	32.4	4.3	2721	US-08-921-195-1	Sequence 1, Appl
12	32.4	4.3	15378	US-08-785-420-1	Sequence 1, Appl
13	31.8	4.2	277	US-09-007-005-3	Sequence 3, Appl
14	31.8	4.2	277	US-09-244-796-3	Sequence 3, Appl
15	31.4	4.2	444	US-09-018-584A-19	Sequence 19, Appl
16	31.4	4.2	444	US-09-018-584A-13	Sequence 13, Appl
17	31	4.1	2088	US-09-351-414-3	Sequence 3, Appl
18	31	4.1	4415	US-09-486-580A-1	Sequence 1, Appl
19	31	4.1	99500	US-09-798-096-10	Sequence 10, Appl
20	30.8	4.1	176373	US-08-220-603A-1	Sequence 1, Appl
21	30.8	4.1	176373	US-09-128-153-17	Sequence 17, Appl
22	30.6	4.1	867	US-08-680-506-1	Sequence 1, Appl
23	30.6	4.1	8321	US-08-971-035-221	Sequence 221, App
24	30.6	4.1	12368	US-08-976-259-20	Sequence 20, Appl
25	30.4	4.0	248	US-09-007-005-32	Sequence 32, Appl
26	30.4	4.0	248	US-09-244-796-32	Sequence 32, Appl
27	30.4	4.0	909	US-09-247-135-131	Sequence 131, App

28	30.4	4.0	17410	1	US-07-841-646-3	Sequence 3, Appl
29	30.4	4.0	17410	1	US-08-147-023-3	Sequence 3, Appl
30	30.4	4.0	17410	1	US-08-447-570-3	Sequence 3, Appl
31	30.4	4.0	17410	2	US-08-449-700-3	Sequence 3, Appl
32	30.4	4.0	17410	2	US-08-449-699A-3	Sequence 3, Appl
33	30.4	4.0	17415	3	US-08-486-343A-1	Sequence 1, Appl
34	30.4	4.0	17415	5	PCT-US95-07329-1	Sequence 1, Appl
35	30.4	4.0	132331	3	US-09-128-135-16	Sequence 16, Appl
36	30.2	4.0	840	1	US-08-690-457-14	Sequence 14, Appl
37	30.2	4.0	840	2	US-08-628-187-14	Sequence 14, Appl
38	30.2	4.0	1702	3	US-08-132-649-3	Sequence 3, Appl
39	30.2	4.0	1702	3	US-08-767-579-3	Sequence 3, Appl
40	30.2	4.0	2080	1	US-08-132-648-1	Sequence 1, Appl
41	30.2	4.0	2080	3	US-08-767-579-1	Sequence 1, Appl
42	30.2	4.0	98844	4	US-09-791-211-10	Sequence 10, Appl
43	29.8	4.0	118	3	US-08-833-167-46	Sequence 46, Appl
44	29.8	4.0	118	4	US-09-344-837A-46	Sequence 46, Appl
45	29.8	4.0	4699	1	US-08-487-890A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-983-502-14
Sequence 14, Application US/08983502
Patent No. 6399327
GENERAL INFORMATION:
APPLICANT: David WALLACH
APPLICANT: Mark P. BOLDIN
APPLICANT: Tanaya M. GONCHAROV
APPLICANT: Yuri V. GOLITSEV
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
TITLE OF INVENTION: AND OTHER PROTEINS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Neimark
SUITE: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/983,502
FILING DATE: 16-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10521
FILING DATE: 14-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 114,615
FILING DATE: 16-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 114,986
FILING DATE: 17-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 115,319
FILING DATE: 14-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 116,588
FILING DATE: 27-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 117,932
FILING DATE: 16-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Broadway, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH=19
TELECOMMUNICATION INFORMATION:

Fri Jul 11 09:23:58 2003

us-09-477-082-2.rst

Page 9

Db 644 YYYYYCCCTTTTTCCTTTCCTTCYCCYCCYYYCCYCCGKRYKSCSTYYTCTTYYCCC 703

QY 588 TTTTATGCTCCACCCCCACCCCTTTCCCTCCCTGCCCCCTGTTTTGTG 639

Db 704 CNYTCGGTCYCCYCCYCCCGCCGCGCGYCGYCCYCTTTCBGRKGTTG 755

```

search completed: July  8, 2003, 05:51:12
job time : 5682.6 secs

```


[illegible]

	Query Match	5.7%	Score 42.8;	DB 17;	Length 1101;	
	Best Local Similarity	48.7%;	Pred. No. 2;			
	Matches	56;	Conservative 19;	Mismatches 40;	Indels 0;	Gaps 0;
OY	552 CCCCTTCGCCCTTCGGCGTGAAGGAGTTCCATTATGTCTTCACACCACCCTC 611 :: Db MCTMTTGTTHTKCMCTCCCCAAATATTTCTCTHTTTTYYYCCCCSSCCCCC 1046					
OY	612 TTCCTTCCTCCCTCTGTTTGTGTGCCCAAAAACAATTCCTTAACGTT 666 :: Db CCCCCCCCCCCCCCTTTTMMMAAAMTWMAAWTATTTT 1101					
	RESULT 13					
	CNS016FT					
	LOCUS					
	DEFINITION	CNS016FT 1201 bp DNA linear GSS 26-JUL-1999				
	ACCESSION	Drosophila melanogaster genome survey sequence Sp6 end of BAC				
	VERSION	BACN15102 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
	SOURCE	AL106691				
	KEYWORDS	AI106691.1 GI:5623268				
	ORGANISM	GSS.				
	REFERENCE	Drosophila melanogaster.				
	AUTHORS	Drosophila melanogaster.				
	TITLE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1201)				
JOURNAL	COMMENT	Direct Submission Submitted (22-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 JURY ENVI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This drosophila melanogaster BAC library (DrosBAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.				
FEATURES	Location/Qualifiers					
SOURCE	1..1201 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_1ld="BACN15102" /clone_1ld="DrosBAC" /plasmid="pBelosBAC11" /note="end : SP6"					
BASE COUNT	272 a 207 c 106 g 438 t 178 others					
ORIGIN						
Query Match	5.7%; Score 42.8;	DB 17;	Length 1201;			
Best Local Similarity	33.1%;	Pred. No. 2;				
Matches	88;	Conservative 54;	Mismatches 124;	Indels 0;	Gaps 0;	
OY	461 TTTTTCCTTCATTCATTCGCGCGGCAGCGTAGTACTTATTCGTGCATTC 520 :: Db TTTTTCCTTCATTCATTCGCGCGGCAGCGTAGTACTTATTCGTGCATTC 841					
OY	521 TGTCGATCACGAATGCCCTGAGGTGACAGACCCCTTCCCTCTTCGCGCTGAAG 580 :: Db AATAAAMHMCRCTWACTTACTTACCTTACCCHCCMCYCCTACCCYCCMYCCYCH 901					
OY	581 GGTTCCCTTATFATCTCCACCCGACCCCTTCGCCCTCGCCGCGCTGTGTTGTCG 640 :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: Db YTTTTCCTTCATTCATTCGCGCGGCAGCGTAGTACTTATTCGTGCATTC 961					
OY	641 CAAAAACAAGTCTCTAAAGCTTTTCGATGTGGATTCGGGAAATTAACGTGCACCC 700 :: Db ATTMCAAHMCCYTCTTTTMMWATTTTYVTMTTCMYWCXHTTHHTTTTTYVCYTATAT 1021					

BASE COUNT	229 a	276 c	89 g	357 t	150 others
ORIGIN	/clone_lib="RRCI-98" /note="end : T7"				

Query Match	5.9%	Score 44.8	DB 17	Length 1101
Best Local Similarity	32.1%	Pred. No. 0.61		
Matches 117	Conservative 66	Mismatches 182	Indels 0	Gaps 0

```
QY      265 TGCATGCCCTGGCCCAAGTCTCCTGTGAGGTTCCTCTCGTAGCCGATGCGTTACATTGCT   324
          :::::::::: : :::: : : ::| | | | | | | | | | | | | | | | | | | | | | |
Db      652 KBBBBSKSTTTGGKKSKIAIYTTTWTHTTTTTTTTTTCYCITTTTTTTTTTTTTCTTTTTT   711
```

Oy 325 ACATTTCACATCIGAGCAGTCTCCAGTTCCTCGGCAACCCTTTGTCTCCAGCTGCC 384
 ||||| || |::: |:: |::||| :|| |:: |
Db 712 TTCTTTTCVTTCTTTTTTTMTCCGCYXXYYTTTTTYTCVYTTTTTTCGCCCCCYCCCTTTC 771

K7 560 ACCGCGCCACCAATGCCTGGTACTCAGGATCCCATTCATCTGTGGGCCGTTTGAGAGAGATCCAGAA 444
 |||::: ::|||:-| | |||:-
Db 772 CCGCCGCCCYCTCCGCCCTCTATTTCATGYCCTCTTTCGCTTTCGCCGCCCTCTTTCCTGCT 831

832 CTTTCCCTCACCYYTTTTCTTCYCCCTCTCYCTTTCTCTCTCTTTTYNCYCCTCCCC 891

[illegible][illegible]

Db 1012 TCCYT 1016

RESOLUTION	AL514103/c	597 bp	mRNA	linear	EST 13-FEB-2000
LOCUS	AL514103				
DEFINITION	AL514103 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA0052F03.3				

ACCESSION AL514103
VERSION AL514103.1 GI:12777597
KEYWORDS EST.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
1 (48,600)	
0.0000000	

AUTHORS	Li, W.B., Gruber, C., Jesse, J. and Polyes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope

FEATURES
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers

```

source
1. .39/
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA0052F03"

```

```

/tissue_type="placenta"
/ote="Vector: pcMWSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

```

cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Peng Liang Life Technologies

Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

BASE COUNT	102 a	70 c	231 g	112 t	82 others
ORIGIN	Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"				

Query Match	5.9%	Score	44.4	DB	9	Length	597
Best Local Similarity	41.5%	Pred	No	0.78			
Matches	78	Conservative	28	Mismatches	82	Indels	0
						Gaps	0

Oy 457 CCACTTTTTCATTTTCATTGGCCGTGGGCGACGGTAAGTACTTTATTCGC 516
| | : : : : :
Db 330 CCCTVTTTTTTTTTTTGSGGGSCGGGSITTTGCCCTVTTTTMMCCCCCCT 271

QY 517 ATTTCGTGGAATACGAAAGCCCTTGAGAGCGCAAGGCCCTTTCCCTATTTCGCCTGG 576
 ::: ::: ::: ::: :::
Db 270 TTTTYYTTTTTCCCCCYYYCYTTTTCSSSSCCYCCCCCBGTTTTTKGCCCCY 211

Db

210 YKGGGKRTTTTTTTTKGGGKKCCCCCCCCCCCCCCCCCCYHMTTTTTTTTTTTTTT 151

Db 150 TTTCMMNV 143

RESULT 11	
CNS0060N	
LOCUS	
CNS0060N	
910 bp	DNA
	linear
	GSS-03-JUN-1999

ACCESSION
VERSION
AL065629.1
GI:19444698
BAC14J21 of RPII-98 library from *Drosophila melanogaster* (fruit
fly), genomic survey sequence.

KEYWORDS
SOURCE
ORGANISM
GSS.
Drosophila melanogaster.
Drosophila melanogaster
Eukaryota: Metazoa: Arthropoda: Hexanoda: Insecta: Pterygota:

REFERENCE
1 (bases 1 to 910)
Genoscope.

Journal of Inequalities and Applications
 Volume 2006 (2006) Genoscope - Centre National de Séquençage :
 Submitted (02-JUN-1999) BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr
 - web : www.genoscope.cns.fr)

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information

melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY.

EcORI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or

FEATURES
source
1. 910
location/Qualifiers
http://bacpac.med.buffalo.edu/drosophila_bac.htm.
filters for hybridization from the BACPAC Resource Center can be
found at

```

/organism="Drosophila
/db_xref="taxon:7227"
/clone="BACR14J21"
/clone_1b="RPC1-98"
metatagset

```

BASE COUNT ORIGIN	202 a	63 c	112 g	198 t	335 others
----------------------	-------	------	-------	-------	------------

Best Local Similarity 18.6%; Pred. No. 16;

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37J10"
/clone_lib="DrosBAC"
/plasmid="PbeloBAC11"
/notes="end : SP6"
BASE COUNT      258 a      174 c      277 g      120 t      272 others
ORIGIN
Query Match      6.2%: Score 46.4; DB 17; Length 1101;
Best Local Similarity 26.0%: Pred. No. 0.24;
Matches 95; Conservative 95; Mismatches 176; Indels 0; Gaps 0;

QY 270 CCTGGCCAGGCTCTCTGTGTTCTCTGACGACGATGCTTGAAGTCTTACTT 329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1085 CCCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 330 TTCACTCTAGAGAGTCTCAGTCTCTGCTACCTTTTGTCTCAAGCTTCCCTGCCG 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1025 YTCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 390 CCTGATGACAGATACAGGACCTCCCTCTCTGAGACCGCTTGAGAGAGAGACTT 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 965 TTTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 906
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 450 TATCAATCCACTTTTCTTTCTTTCTTTCTTTGAGCCCTGGGGGCGAGGTTAAGTCTTA 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 905 YTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 846
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 TTTCTGATCTGTGGAATACAGAAAGCCCTGAGTGTGACAGACCGCTTCCCTCTTCG 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 786
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 570 GCTCCTGAAGGGGTTCTTTATGTTCTTCACACCCCAACCTTTCCCTCCCTGCTCT 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 785 CTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 630 GTTTT 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 725 TCTTTT 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
CNS0181N/c 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN37J10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL108773.1 GI:5629077
VERSION AL108773
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
          Genoscope.
          Direct Submission
          Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seget@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the European Drosophila Genome Project (BDGP) -
          http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
          library (Dros BAC) was made by Alain Billaud at CPH (Centre
          d'Etude du polymorphisme Humain) with funding provided by a MRC
          project grant. The DNA was prepared from embryos by Alain Bucheton
          and Genevieve Payan. It has been constructed in the vector
          PbeloBAC11.
FEATURES
Source 1..1101
          location/Qualifiers
          /organism="Drosophila melanogaster"

```

```

/db_xref="taxon:7227"
/clone="BACN37J10"
/clone_lib="DrosBAC"
/plasmid="PbeloBAC11"
/notes="end : SP6"
BASE COUNT      125 a      364 c      197 g      116 t      299 others
ORIGIN
Query Match      6.1%: Score 46.2; DB 17; Length 1101;
Best Local Similarity 21.5%: Pred. No. 0.27;
Matches 59; Conservative 99; Mismatches 117; Indels 0; Gaps 0;

QY 452 TCAATCCACTTTTCTTTCTTTCTTTCTTTGAGCCCTGGGGCGAGGTTAAGTCTTAT 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1028 YCCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 969
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 512 CTGCTATCTGTGGAATACAGATGCTCCCTGAGGTGACAGACCGCTTCCCTCTTCGG 571
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 968 CCCCCTCCCTCCCTCTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 909
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 572 TCTGAGGGGGTCTCTTTATGTTCTTCACACCCCAACCTTTCCCTCCCTGCTCT 631
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 908 CCCCCCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 849
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 632 TTTTGTGGCCCAAAACAGTCTCTAAAGTTTGTGATGTGATGCGGGAATTA 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 848 YTTCTCTCTCTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 789
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 CCTGACCCCTTTCGAATGACCTTTTCTTTT 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 788 CTTCTCTCTCTTTCTTTCTTTCTTTCTTTT 754
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
CNS0016Z 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR24F06 of RPCR-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL068148
VERSION AL068148
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
          Genoscope.
          Direct Submission
          Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
          BP 191 91006 EVRY cedex - FRANCE (E-mail : seget@genoscope.cns.fr
          - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org. The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osogawa and
          Aaron Mammeter in Pieter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCR-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          isogenic strain Y2; cn bw sp. The same strain used for the BDGP's
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Source 1..1101
          location/Qualifiers
          /organism="Drosophila melanogaster"
          /db_xref="taxon:7227"
          /clone="BACR24F06"

```



```

QY 73 CGTGAGTTAGGACAGTATGAGGACTCG---GAGACTGCATGTGTCGACGAAAGGCTGG 129
DB 262 CGTGAGTAAGGACAGTATGAGGAGACTGCGAGAGCTGCGAGTGGTGCCAGAAAGGCTGG 321
QY 130 AGCGGGTGAAGTGCCTGTGTGCC 150
DB 322 AGCGGATATATATCTCTCGCC 342

RESULT 4
CNS005TE 997 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767.1 GI:4943573
VERSION AL060767.1
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source
1..997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 89 a 99 c 13 g 258 t 538 others
ORIGIN
Query Match 7.8%; Score 58.6; DB 17; Length 997;
Best Local Similarity 17.1%; Pred. No. 0.00018;
Matches 101; Conservative 185; Mismatches 305; Indels 0; Gaps 0;

QY 163 TCACAGCAACCAATATTTTGTCTGACTGTCTGAGAAGAGGAGGCTGGGG 222
DB 386 TANNANNAANAAAAAANNNTAGNNNNNNNNNNNNNNAANANNANNANNANGCTN 445
QY 223 TGGGAGCAACTGATCGCCTCTGAGAGAACCTCTGCTGCTGGCCAGGTC 282
DB 446 NNNNNNNNNNNNNNNNANAYYYTCTYTYTYYTYYTYYTYYTYYCTYYCCYCY 505
QY 283 TCGTGTGGTTCCTCTGAGCGAGCTTGTGACTTGTACTTTTTCACCTGAGCA 342
DB 506 YTTCCYYYYYTYYTYYTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 565
QY 343 GTCTCCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 402

```

```

DB 566 YTTTCYYYTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 625
QY 403 TACAGGACTCCCTCTCTGTGAGCCGTTTGGAGATCCAGACATTATCAATCAGCT 462
DB 626 CYCYCYYYTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 685
QY 463 TTTTCTTTTTCATTTGGCCCTGGGGCCGAGAGGTTAATCTATCTGATCTG 522
DB 686 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 745
QY 523 TCAGATCAGCAATGCCCTGAGTGCACAGCCCTTCCCTCTTTCGCGCTAGAGGG 582
DB 746 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 805
QY 583 TTTCTCTTTATGCTTTCACCCACCCCTTCCCTCTCTCTCTCTCTCTCTCTCT 642
DB 806 TTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 865
QY 643 AAAAACAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 702
DB 866 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 925
QY 703 TTGCAAAATGACTTTTCTTTTGTGATCTGTACCTGTTTAACTT 753
DB 926 YTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 976

RESULT 5
CNS00418 987 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR09C16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL066537.1 GI:4942778
VERSION AL066537.1
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 987)
Genoscope.
REFERENCE Direct Submission
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Source
1..987
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="BACR09C16"
/note="end : TET3"

BASE COUNT 238 a 162 c 17 g 177 t 393 others
ORIGIN
Query Match 7.4%; Score 55.4; DB 17; Length 987;

```

FEATURES
Source
Location/Qualifiers
1. 522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp313F1234"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTRiPLEX2; Site_1: SfiI; Site_2: SfiI; B; cDNA-collection"

BASE COUNT 120 a 116 c 133 g 149 t 4 others

Query Match 16.6%; Score 125.2; DB 9; Length 522;
Best Local Similarity 94.2%; Pred. No. 1.4e-21;
Matches 130; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 72
|||||
DB 226 TATTGAAGTAAAGAACTTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 285
|||||

QY 73 CGTGAAGTTAGGACGTTAGGAGCTCGAGACTCGATGTGTGCCAGAAAGGTTGAGC 132
|||||
DB 286 CGTGAAGTTAGGACGTTAGGAGCTCGAGACTCGATGTGTGCCAGAAAGGTTGAGC 345
|||||

QY 133 GGGTGAAGTGGCTGTGCC 150
|||||
DB 346 GGATTATATCTCTCTGCC 363
|||||

RESULT 2
AL702116 421 bp mRNA linear EST 22-MAR-2002
LOCUS DKFZp686E18155.r1.686 (synonym: hicc3) Homo sapiens cDNA clone
AL702116 DKFZp686E18155.5, mRNA sequence.
ACCESSION AL702116.1 GI:19685471
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 421)
AUTHORS Bioecker, H., Boecker, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.
TITLE EST (Bioecker, H., Boecker, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Bioecker H
MIPS
Am Klopferpfitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.
No si sequence available.
This clone (DKFZp686E18155) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686E18155"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTRiPLEX2; Site_1: SfiI; Site_2: SfiI; B; cDNA-collection"

BASE COUNT 95 a 89 c 113 g 124 t

ORIGIN
Query Match 16.4%; Score 123.6; DB 9; Length 421;
Best Local Similarity 93.5%; Pred. No. 3.7e-21;
Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 72
|||||
DB 207 TATTGAAGTAAAGAACTTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 266
|||||

QY 73 CGTGAAGTTAGGACGTTAGGAGCTCGAGACTCGATGTGTGCCAGAAAGGTTGAGC 132
|||||
DB 267 CGTGAAGTTAGGACGTTAGGAGCTCGAGACTCGATGTGTGCCAGAAAGGTTGAGC 326
|||||

QY 133 GGGTGAAGTGGCTGTGCC 150
|||||
DB 327 GGATTATATCTCTCTGCC 344
|||||

RESULT 3
B1824550 804 bp mRNA linear EST 04-OCT-2001
LOCUS 603035156F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176548.5',
DEFINITION mRNA sequence.
B1824550
ACCESSION B1824550.1 GI:15936100
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 804)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11439 row: 0 column: 13
High quality sequence stop: 620.
Location/Qualifiers
1. 804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5176548"
/clone_11b="NIH_MGC_115"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-OT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."

BASE COUNT 226 a 187 c 209 g 182 t

ORIGIN
Query Match 12.8%; Score 96.2; DB 13; Length 804;
Best Local Similarity 85.1%; Pred. No. 4e-14;
Matches 120; Conservative 0; Mismatches 18; Indels 3; Gaps 1;

QY 13 TATTGAAGTAAAGAACTTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 72
|||||
DB 202 TATTGAAGTAAAGAACTTCTCTGGAGCCTTCCACCCCTTCCCTGCTGAGCA 261
|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 ; Search time 5677.6 Seconds
(without alignments)
2147.949 Million cell updates/sec

Title: US-09-477-082-2

Perfect score: 753
Sequence: 1 aatgaccgcgtattgaa.....taccctgtttttaccctt 753

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST :
1: em_estba : *
2: em_estchum : *
3: em_estcin : *
4: em_estimu : *
5: em_estmuv : *
6: em_estovl : *
7: em_estro : *
8: em_hic : *
9: gb_estcl : *
10: gb_estl2 : *
11: gb_hic : *
12: gb_est3 : *
13: gb_est4 : *
14: gb_est5 : *
15: em_estfun : *
16: em_estom : *
17: gb_gss : *
18: em_gss_hum : *
19: em_gss_inv : *
20: em_gss_pln : *
21: em_gss_vrt : *
22: em_gss_fun : *
23: em_gss_mam : *
24: em_gss_mus : *
25: em_gss_other : *
26: em_gss_pro : *
27: em_gss_rod : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	125.2	16.6	522	9	AL600571
2	123.6	16.4	421	9	AL702116
3	96.2	12.8	804	13	BI824550
4	58.6	7.8	987	17	CNS0057E
5	55.4	7.4	987	17	CNS00418
6	46.8	6.2	772	17	CNS02BMV

Result No.	Score	Query Match	Length	DB ID	Description
7	46.4	6.2	1101	17	CNS017RP
8	46.2	6.1	1101	17	CNS0181N
9	44.8	5.9	1101	17	CNS0016Z
10	44.4	5.5	597	9	AL514103
11	43.2	5.7	910	17	CNS0060N
12	42.8	5.7	1101	17	CNS0160T
13	42.8	5.7	1201	17	CNS016FT
14	42.4	5.6	571	17	AO646132
15	42.2	5.6	846	17	CNS00KOY
16	42.2	5.6	1101	17	CNS00Z4D
17	42.2	5.6	1143	17	BI1483
18	41.8	5.5	1124	17	BI1278
19	41.2	5.5	1100	17	CNS00FSE
20	41.1	5.4	362	9	AI349834
21	41.1	5.4	739	17	BI179653
22	41.1	5.4	739	17	CNS071OR
23	41.1	5.4	1100	17	CNS00DER
24	41.1	5.4	1107	17	AG129299
25	40.8	5.4	262	13	BM119154
26	40.8	5.4	1061	17	CNS001FJ
27	40.8	5.4	1201	17	CNS0160T
28	40.6	5.4	914	17	AG184648
29	40.4	5.4	980	17	CNS03844
30	40.2	5.3	934	17	CNS00D4F
31	40.2	5.3	1117	17	CNS00Z2S
32	40.2	5.3	625	10	AM053904
33	40.0	5.3	1101	17	CNS00GC9
34	40.0	5.3	1128	13	BI956472
35	39.8	5.3	439	17	BH749911
36	39.8	5.3	1101	17	CNS0030C
37	39.8	5.3	1220	14	BQ058373
38	39.8	5.3	1352	12	BE896162
39	39.8	5.3	1364	13	BM474596
40	39.6	5.3	377	17	AO820430
41	39.6	5.3	1201	17	CNS016F2
42	39.4	5.2	693	17	AG043041
43	38.4	5.2	919	17	CNS006S5
44	38.4	5.2	1201	17	CNS0010J
45	39.2	5.2	680	17	AG102143

ALIGNMENTS

RESULT 1
LOCUS AL600571 522 bp mRNA linear EST 14-AUG-2001
DEFINITION DKFZP313F1234.r1 313 (synonym: hicc2) Homo sapiens CDNA clone
ACCESSION DKFZP313F1234.5, mRNA sequence.
VERSION AL600571
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Ansoerge, W., Winkner, U., Mewes, W., Weil, B. and Wiemann, S.
TITLE EST (Ansoerge, W., Winkner, U., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansoerge W

MIPS
Am Klopferstutz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the CDNA sequencing consortium of the
German Genome Project.
No 5' sequence available.
This clone (DKFZP313F1234) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

DR P-PSDB; ABB59675.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 5816; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SO Sequence 7723 BP; 2272 A; 1661 C; 1868 G; 1902 T; 0 other:
 Query Match 4.9%; Score 37.2; DB 23; Length 7723;
 Best Local Similarity 49.5%; Pred. No. 1.9;
 Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
 QY 280 GTCTCCGTGTGGTTCTCTCTGAGCGGATGCTTGACTTGCTACTTTTTCAGCTGTA 339
 DB 2688 GCGCTTCTTCCCGCTTTTGTTCCTCACTCAATAGCTTGCGCTTTTTCGCTTA 2629
 QY 340 GCACTCTCAGTTCCTCTGCTACCTTTTGTCTCCAGCTTCCGCGCTCGAATGC 399
 DB 2628 ACCGCTCCAACTCTCTGCGCAGAGATTAGATCAGACTTGCGCTTGCTGACTTT 2569
 QY 400 AGATACAGGACACCCCTCTGTGAGACCGTTTGAGAGTCCAGAAAGCTTATCAATCA 459
 DB 2668 GCACCGCTTCTTACCGGCGACAAATTAATTTGTGAGAGTGTGCTTGCGCTCA 2509
 QY 460 CTTTCTTTCTTTT 473
 DB 2508 GGTTTTATCTT 2495
 RESULT 15
 AAA10594/c
 ID AAA10594 standard; DNA; 10732 BP.
 AC AAA10594;
 XX 29-JUN-2000 (first entry)
 DT Gene encoding a subunit of cellulose synthase.
 DE Cellulose synthase; cellulose production; increase yield; ds.
 KW *Vigna angularis*.
 OS *Vigna angularis*.
 XX JP2000060568-A.
 PN 29-FEB-2000.
 PD 26-AUG-1998; 98JP-0239998.
 PF 26-AUG-1998; 98JP-0239998.
 PR (MIZU/) MIZUNO K.
 PA (OJIP) OJI PAPER CO.
 XX WPI: 2000-342371/30.
 DR P-PSDB; AAV85179.
 XX A gene encoding a cellulose synthetic equipment - for the improvement
 PT in the amount of cellulose synthesised in a plant body

XX
 PS Claim 2; Page 14-21; 32pp; Japanese.
 XX
 CC This sequence represents a gene encoding a subunit of the cellulose
 CC synthase complex of *Vigna angularis*. The invention relates to subunits of
 CC cellulose synthetic equipment, that can be used to increase the amount of
 CC cellulose synthesised by a plant. The proteins and genes encoding them
 CC can also be used to improve the properties of the cellulose being
 CC produced by a plant.
 XX
 SO Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other:
 Query Match 4.8%; Score 36.4; DB 21; Length 10732;
 Best Local Similarity 18.3%; Pred. No. 3.9;
 Matches 91; Conservative 167; Mismatches 237; Indels 3; Gaps 1;
 QY 146 TTGCCAAGTGGCTCTTCCACAGGAAACCAATATTTTGTGTTCTTGACTGCTAG 205
 DB 9510 BYSDBRCSBDSCTCNCYSCDSBSRYSTYDACYTDAKTBCTYSDNCCNSRKYNST 9451
 QY 206 AACAAGGCGTGGGGGTGGGGAAGCACTTGATCTGCCCTTGAGAGACACCTGTGT 265
 DB 9450 YGCSRTBSRNSRSTCCBTBTSRGNCCIDAYDANSTRIDYDACYSTDASTBTSIST 9391
 QY 266 GCTGCTGCGCCGAGGCTCCTGTGTTCTCTCTGAGCGGATGCTTGACTTGGCTA 325
 DB 9390 BYSNSRYDAYSSRYSCYTYCTDYSSTCTRCANKCTBCNSRRAKSTTTTTRCYCT 9331
 QY 326 CTTTTCACCTGAGCAGTCTCAGTCTCTGCTACCTTTTGTGCTCCAGCTCCCT 385
 DB 9330 SRCYTBCTCYTNCRSRGSRTSRSDAYSCNCSRSRSTBCTYDACYCYYS 9271
 QY 386 GCGGCTCATGACAGATACAGGACGACCTCTGTGACCGCTTGAGAGAGTCCAGAG 445
 DB 9270 DCCYTCNSTCYTCCSRCTBCTTBCYRAYDYSDANSTRRAKSCCTTYAKTTR 9211
 QY 446 ACT--TTATCAATCCACTTTTCTTTTCAATTTGAGCCCTGGGGCCGACGTTAG 502
 DB 9210 CSDSTCCTTNCYDANCCYCTSTNCSNCTCTTDDYSSRNSTSDYNSSTTBTCT 9151
 QY 503 TACTTATTCGTGCAATCTGTGCAATGCAAGATGCGCGAGGTGACAGCCCTTCC 562
 DB 9150 TCTYSRYSDACYSNOCCTBAKCYTOSRDCSRVSTTSRAKBTSTSDCTNCCNCCC 9091
 QY 563 TCTTTCGCTGCTGAGGGGTTCTTTATGCTCTTCCACCCCACTTCCCTCCCT 622
 DB 9090 YTTBDCYTNCAKSRGSRTRTCSHSDTTTBCYTCVATBCYATNCCNCTSTNCSR 9031
 QY 623 GCGCTCTGTTTGTGTC 640
 DB 9030 AKYSTBCTTCTTSRSRRC 9013

Search completed: July 8, 2003, 03:11:10
 Job time : 755.255 secs

PR 07-OCT-1999; 99US-0156029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159684.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.0%; Score 37.4; DB 21; Length 1522;
Best Local Similarity 62.1%; Pred. No. 0.72;

Matches 59; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 281 TCTCCGTCGTCGTCCTCTCTGAGCCGATGCTTGGACTTGTCTTTCACCTGAG 340
DB 917 TCTCCTTTACAGTTCCGCTTCGCGTTCCTCTGACTTGATATTTCTTTCTCCG 858
QY 341 CAGTCTCAGTTCCTGCTGACCTTTTGTCTCTCC 375
DB 857 CCATCTCTTTTACAGCTACTTCTGCTCTCC 823

RESULT 13

ABL03779/c
ID ABL03779 standard; cDNA: 4983 BP.

AC ABL03779;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5819.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PERE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.
DR P-PSDB; ABB59676.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 5819; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB5737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4983 BP; 1500 A; 1137 C; 1257 G; 1089 T; 0 other;

Query Match 4.9%; Score 37.2; DB 23; Length 4983;
Best Local Similarity 49.5%; Pred. No. 1.5;

Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 280 GTCTCCTGTGTCGTCCTCTCTGAGCCGATGCTTGGACTTGTCTTTCACCTGTA 339
DB 1196 GCCTTCCTTCCGCTTTGTTTCCACACATATAGCTTGGCTTTTTCGCGTTA 1137
QY 340 CCACTCTCAGTTCCTCTGCTACCTTTTGTCTCCAGCTTCCCTCCGCTGGAATGC 399
DB 1136 ACCGCTCCCAATTCCTCTGCGAGATTTAGCACTGAGCTTGGCTTGGCTTGAATTT 1077
QY 400 AGATACCGAGACCTCTCTGAGCCGATGCTTGGAGAGTTCAGAGACTTATCAATCCA 459
DB 1076 GCAGCCGTGCTTACCGCGCACAATATATTTGTTGAGAGTGATTCCTTTCGCCCTCA 1017
QY 460 CTTTCTTTCTTTT 473
DB 1016 GGTCTTTTAACTT 1003

RESULT 14

ABL03778/c
ID ABL03778 standard; cDNA: 7723 BP.

AC ABL03778;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5816.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

XX WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PERE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128214.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139484.
PR 18-JUN-1999; 99US-0139485.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142305.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147316.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151103.
PR 31-AUG-1999; 99US-0151138.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

```

PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144682.
PR 20-JUL-1999; 9905-0144864.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148665.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.

```

```

PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 13-OCT-1999; 9905-0159299.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161922.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

```

```

Query Match          5 0%; Score 37.4; DB 21; Length 1227;
Best Local Similarity 62.1%; Pred. No. 0.65;
Matches 59; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```

```

QY 281 TCTCTGTGTCGTTCTCTCTGAGCGATGCTTGACTTTCCTACTTTCACCTCTGAG 340
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 919 TCTCTTTTCAGCTTCCGCTTCGCGTCCCTCGACTTTCATATTTCTTCTCTCG 860
QY 341 CAGTCTCCAGTTCCTCTGCTGCTACCTTTTGTCTCTCC 375
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 859 CCATCTCTTTTTCAGCTACTCTCTGCTCTCC 825

```

RESULT 12

AAC51060/c
ID AAC51060 standard; DNA; 1522 BP.

XX AAC51060;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 67125.

XX Hybridisation assay; genetic mapping; gene expression control;

KW Protein identification; signal transduction pathway;

KW metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EPI03405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.

PR 09-MAR-1999; 9905-0123548.


```

FT      /number="2"
FT      1079..1156
FT      /tag= e
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      1080..1083
FT      /tag= f
FT      1565..1599
FT      /tag= g
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      1566..1567
FT      /tag= h
FT      1625..1919
FT      /tag= i
FT      /number="3"
FT      1920..2511
FT      /tag= j
FT      /number="3"
FT      2512..2829
FT      /tag= k
FT      /number="4"
FT      2830..4471
FT      /tag= l
FT      /number="4"
FT      3088..3138
FT      /tag= m
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      3088..3091
FT      /tag= n
FT      3779..3823
FT      /tag= o
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      3779..3780
FT      /tag= p
FT      3874..3901
FT      /tag= q
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      3875..3878
FT      /tag= r
FT      4017..4272
FT      /tag= s
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      4017..4022
FT      /tag= t
FT      4472..4645
FT      /tag= u
FT      /number="5"
FT      4646..5385
FT      /tag= v
FT      /number="5"
FT      4914..4982
FT      /tag= w
FT      /standard_name="microsatellite"
FT      repeat_unit
FT      4915..4916
FT      /tag= x
FT      5356..5484
FT      /tag= y
FT      /number="6"
FT      5485..6897
FT      /tag= z
FT      /number="6"
FT      6898..7066
FT      /tag= aa
FT      /number="7"
FT      7067..7404
FT      /tag= ab
FT      /number="7"
FT      7405..8300
FT      /tag= ac
FT      /number="8"
FT      8301..8300
FT      /tag= ad
FT      /number="8"

```

```

XX      02-OCT-2000; 2000WO-GB03784.
XX      PF
XX      01-OCT-1999; 99ES-0002169.
XX      PR
XX      (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
XX      (KIDD/) KIDDLE S.
XX      PA
XX      Blanco DL, Bernad Miana A, Dominguez Lopez O, Garcia Diaz M;
XX      WPI: 2001-335565/35.
XX      DR
XX      P-PSDB; AAB97498.
XX      DR
XX      Novel isolated DNA polymerase lambda polypeptide useful for treating
XX      pr neurodegenerative disorders, cancer, graft rejection, arthritis,
XX      psoriasis, and for repairing or reconstructing DNA molecule
XX      PT
XX      Disclosure: Page 94-101; 121pp; English.
XX      PS
XX      The present invention provides the protein and coding sequences of human
XX      CC and murine DNA polymerase lambda (Pol lambda). This protein is thought to
XX      CC be involved in DNA repair. The sequences can be used in the diagnosis and
XX      CC treatment of cancer, neurodegenerative diseases, immunosuppression,
XX      CC psoriasis, arthritis and graft rejection. The human gene is found at
XX      CC chromosome 10q24, from which several isoforms are produced, and the
XX      CC murine gene is found on chromosome 19. The present sequence is the murine
XX      CC Pol lambda gene. Pol lambda is also designated POLL and Pol kappa.
XX      CC
XX      Sequence 8300 BP; 1881 A; 2296 C; 2189 G; 1934 T; 0 other;
XX      SQ
XX      Query Match 5.1%; Score 38.4; DB 22; Length 8300;
XX      Best Local Similarity 50.0%; Pred. No. 0.82;
XX      Matches 96; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
XX      QY 183 TTTTGTCTTCTGACTTGTCTGCTGAGAACAGGCGCTGTGGGGGTGGGAGACACTGTGATCT 242
XX      ||||| || || ||||| || || ||||| || || ||||| || || ||||| || || |||||
XX      Db 3937 TTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 3996
XX      QY 243 GCCCTCTGAGGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
XX      || || || || || || || || || || || || || || || || || || || || || || ||
XX      Db 3997 AACCTCAGACATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 4056
XX      QY 303 AGCCGATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 362
XX      ||| ||||| || || || || || || || || || || || || || || || || || || || ||
XX      Db 4057 TGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTC 4116
XX      QY 363 CTTTGTGCTCT 374
XX      || || || || || || || || || || || || || || || || || || || || || || ||
XX      Db 4117 TGCTCTGCTCTC 4128
XX      QY
XX      Db
XX      RESULT 10
XX      AA193505
XX      ID AA193505 standard; CDNA; 439 BP.
XX      AC
XX      AA193505;
XX      AC
XX      06-NOV-2001 (first entry)
XX      DE Human polynucleotide SEQ ID NO 13565.
XX      XX
XX      Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX      KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX      KW tissue growth factor; immunomodulatory; cancer; leukemia;
XX      KW nervous system disorders; arthritis; inflammation; ss.
XX      OS Homo sapiens.
XX      OS
XX      PN WO200164835-A2.
XX      PD 07-SEP-2001.
XX      PF 26-FEB-2001; 2001WO-US04927.

```

PR 28-JUL-2000; 2000US-221607P.
 PR 02-MAY-2001; 2001US-287724P.
 XX
 XX (COMP-) COMPUGEN INC.
 XX
 XX Shoshan A, Masserman A, Mintz E, Mintz L, Faigler S;
 XX WPI; 2002-257383/30.
 XX
 PT New oligonucleotide libraries comprising oligonucleotides which
 PT selectively hybridize to mRNAs transcribed from a transcription unit of
 PT a genome, useful for detecting tissue-, pathology-, and
 PT developmental-specific genes
 XX
 PS Example 1; SEQ ID 31650; 47pp; English.
 XX
 CC The present invention describes oligonucleotide libraries for detecting
 CC messenger RNAs that populate a (sub-)transcriptome, where the
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
 CC transcription units that populate a genome. The library comprises
 CC several oligonucleotides, each capable of hybridizing selectively to a
 CC set of messenger RNAs transcribed from a given transcription unit of
 CC the genome, which encodes one or more messenger RNA splice variants.
 CC The oligonucleotide libraries are useful for detecting mRNAs from a
 CC biological sample, in expression profiling studies, in qualitatively or
 CC quantitatively characterizing the corresponding transcriptome, and in
 CC detecting RNA transcripts and splice variants of human or animal
 CC transcriptomes. The libraries may also be used as specialized mini
 CC libraries to detect transcripts of a sub-transcriptome under a
 CC particular biological or pathological state, and so allowing the
 CC detection of tissue- and pathology-specific genes such as those genes
 CC only expressed in specific tissue under a specific pathological
 CC condition; to detect developmental specific genes; and to detect RNA
 CC transcripts and splice variants of a transcriptome of a patient suffering
 CC from a particular disorder. ABN27253 to ABN59589 represent
 CC oligonucleotide sequences from rats, humans and mice, which are used in
 CC the exemplification of the present invention.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 60 BP; 14 A; 8 C; 28 G; 10 T; 0 other;
 XX
 Query Match 8.0%; Score 60; DB 24; Length 60;
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 72 ACGTGAGATTAGCGAGTTAGGGAGCTCGAGACTGCGATGTCGCCAGAAAGGTCGAG 131
 Db 1 ACGTGAGATTAGCGAGTTAGGGAGCTCGAGACTGCGATGTCGCCAGAAAGGTCGAG 60
 XX
 RESULT 8
 ABL3323/C
 ID ABL3323 standard; DNA; 9741 BP.
 XX
 AC ABL3323;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1296.
 XX
 KW Human, immune system disease; cytosine methylation; antiasthmatic;
 KW antiatherosclerotic; antiaiemic; cytostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX

PN WO200200928-A2.
 XX
 XX 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 DR Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 1296; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SQ Sequence 9741 BP; 2623 A; 195 C; 1863 G; 5060 T; 0 other;
 XX
 Query Match 5.2%; Score 39.4; DB 24; Length 9741;
 Best Local Similarity 59.3%; Pred. No. 0.44;
 Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
 QY 640 CCCAANAACAGCTCTTAACGTTTCGATGCGATCCGGAAATTAACCTGCACC 699
 Db 662 CCTAANAACAGCTCTTAACGTTTCGATGCGATCCGGAAATTAACCTGCCTA 603
 QY 700 CGTTGCAAAATGAACTTTTTCCTGATCCTGACACTGGTTTAACT 752
 Db 602 TTTATAAATACGAAATTTTCACTTTTTCCTGATCCTGACACTGGTTTAACT 550
 XX
 RESULT 9
 AAH25711
 ID AAH25711 standard; DNA; 8300 BP.
 XX
 AC AAH25711;
 XX
 DT 14-AUG-2001 (first entry)
 XX
 DE Murine DNA polymerase lambda gene.
 XX
 KW Mouse; human; DNA polymerase lambda; DNA repair; isoform; cancer;
 KW neurodegenerative disease; gene therapy; pol kappa; POLT; chromosome 19;
 KW aging; immunosuppression; psoriasis; arthritis; graft rejection; ds.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT 1..58
 FT /tag= a
 FT /number= "1"
 FT Intron
 FT 59..827
 FT /tag= b
 FT /number= "1"
 FT Exon
 FT 828..964
 FT /tag= c
 FT /number= "2"
 FT Intron
 FT 965..1624
 FT /tag= d

CC DNA chip analysis as given in the specification, and comparing
CC the expression level to an expression level in an unactivated

PF 20-JUL-2001; 2001WO-IB01903.
XX

20-JUL-2001; 2001WO-1B01903.

ID AAT61405 standard; cDNA; 2887 BP.
 XX AAT61405:
 XX
 XX
 DT 29-OCT-1997 (first entry)
 DE MACH isoform alpha1 coding sequence.
 XX
 KM MACH: MORT-1 binding protein; mediator of receptor toxicity; cell death;
 KM antibody; FAS ligand receptor; FAS-R; death domain region; septic shock;
 KM tumour necrosis factor; tumour; HIV-infection; oligodendrocyte death;
 KM apoptosis/programmed cell death; p55-R; graft rejection; acute hepatitis;
 KM autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide;
 KM TNF; therapy: ss.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 291..1731
 FT /tag= a
 FT /product= MORT-1 binding protein (MACH) isoform alpha1
 XX
 PN MO970398-A1.
 XX
 PD 06-FEB-1997.
 PF 14-JUN-1996; 96MO-US10521.
 XX
 PR 16-APR-1996; 96IL-0117932.
 PR 16-JUL-1995; 95IL-0114615.
 PR 17-AUG-1995; 95IL-0114986.
 PR 14-SEP-1995; 95IL-0115319.
 PR 27-SEP-1995; 95IL-0116588.
 XX
 PA (WEIN/) WEINMURZEL H.
 PA (YEDA) YEDA RES & DEV CO LTD.
 PI Boldin M, Goltssev YV, Goncharov T, Wallach D:
 DR MPI: 1997-132570/12.
 DR P-PSDB; AAM11891.
 XX
 PT New DNA encoding MACH protein that interacts with MORT-1 protein -
 PT to mediate intracellular effects of FAS or TNF receptors, partic.
 PT for regulating apoptosis in tumours, virus-infected cells etc.
 XX
 PS Example 3; Page 116-117; 163pp; English.
 XX
 CC AAT61396, and AAT61405-T61411 represent coding sequences for different
 CC isoforms of MACH. MACH is a binding protein for the mediator of receptor
 CC toxicity (MORT-1) protein. MORT-1 binds to the FAS ligand receptor
 CC (FAS-R) death domain region, and triggers part of the cell death
 CC signalling cascade in mammalian cells. Vectors containing these
 CC sequences, the encoded proteins, and antibodies (Ab) against them are
 CC used to modulate the effect of FAS-R ligand or TNF on cells that carry
 CC FAS-R or p55-R. This is specifically for treating tumours, HIV-infected
 CC cells or other diseased cells, by control of apoptosis/programmed cell
 CC death. The encoded proteins are mediators of the cell death pathway
 CC initiated by TNF and FAS-R binding, i.e. it mimics or enhances the
 CC effect of MORT-1 where increased cytotoxicity is required. To inhibit the
 CC effect of MORT-1, e.g. in cases of septic shock, graft rejection and
 CC acute hepatitis, sequences encoding antisense molecules or ribozymes, or
 CC Ab against the protein, are used. Compounds that inhibit MACH are
 CC potentially useful for controlling MACH activity e.g. in cases of
 CC autoimmune disease, oligodendrocyte death in multiple sclerosis or
 CC AIDS-inhibited T-cell suicide. The MACH proteins can also be used to
 CC isolate and characterise other proteins and receptors involved in
 CC signalling and for Ab production. The Ab can be used to purify the new
 CC proteins and for diagnosis of conditions involving abnormal function of
 CC FAS-R mediated cellular effects.
 XX
 SQ Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;

Query Match 16.4%; Score 123.6; DB 18; Length 2887;
 Best Local Similarity 93.5%; Pred. No. 2.8e-27;
 Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 QY 13 TATTGAAGTAAAGAACTTCTTCCTGGAGACCTTTCCACCCCTTCCCTGTGAGCA 72
 DB 144 TATTGAAGTAAAGAACTTCTTCCTGGAGACCTTTCCACCCCTTCCCTGTGAGCA 203
 QY 73 CGTGAGTTAGGACAGTGGGAGCTGGAGACTGGATGGTGGCCAGAGAAAGGTGAGC 132
 DB 204 CATTGAGTTAGGACAGTGGGAGCTGGAGACTGGATGGTGGCCAGAGAAAGGTGAGC 263
 QY 133 GGGTGAAGTGGCTGTGACC 150
 DB 264 GGATTATATTCCTGCTGCC 281
 RESULT 5
 AAD03915
 ID AAD03915 standard; DNA; 2887 BP.
 XX
 AC AAD03915;
 XX
 DT 02-JUL-2001 (first entry)
 DE Human caspase-8 DNA.
 XX
 KM Human: caspase-8; MACH; FLICE; Mch3;
 KM cysteine protease; apoptosis; caspase expression cassette; metastasis;
 KM tumour; cathepsin B; urokinase; proliferation; gene therapy; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 292..1731
 FT /tag= a
 FT /product= "Human caspase-8"
 XX
 PN WO200129232-A2.
 PD 26-APR-2001.
 XX
 PR 19-OCT-2000; 2000MO-US28941.
 PR 20-OCT-1999; 99US-0160559.
 PR 14-AUG-2000; 2000US-0225564.
 XX
 PA (SCIO-) SCIOS INC.
 PI Cordell B, Li Y;
 DR MPI; 2001-290920/30.
 DR P-PSDB; AAE00605.
 XX
 PT Novel fusion polypeptide comprising first and second caspase subunit
 PT separated by cleavage site not associated in nature with caspase
 PT subunit, useful for cloning gene encoding enzymes involved in
 PT proteolytic cleavage
 XX
 PS Disclosure: Fig 15; 116pp; English.
 XX
 CC The present sequence is a DNA encoding human Caspase-8 also known as
 CC MACH, FLICE and Mch3. Caspases are a family of cysteine proteases, that
 CC participate in the initiation and execution of apoptosis. Caspases exist
 CC as pro-enzymes, activated by cleavage into a large and small subunit,
 CC occurring after specific aspartic acid residues within the pro-enzyme
 CC sequence. The present invention relates to a method for functional
 CC cloning of genes encoding proteins or enzymes involved in proteolytic
 CC cleavage. The invention is based on the use of caspase expression
 CC cassettes comprising the coding sequence of a proteolytic cleavage site
 CC flanked by sequences encoding two caspase subunits. A fusion polypeptide
 CC comprising a first and a second caspase subunit, separated by a cleavage
 CC site not associated in nature, is useful for cloning gene encoding

20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 08-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 PS
 PS Disclosure; SEQ ID NO 32029; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK6784 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 45017 BP; 12919 A; 10037 C; 9566 G; 12495 T; 0 other;

Query Match 70.4%; Score 529.8; DB 22; Length 45017;
 Best Local Similarity 91.68; Pred. No. 9, se-152;
 Matches 675; Conservative 0; Mismatches 42; Indels 20; Gaps 10;

QY	13	TATTGAAGTAAAGAACTCTTCCGAGAGCCTTCCACCCCTTCCCTGCTGAGCA	72
DB	25499	TATTGAAGTAAAGAACTCTTCCGAGAGCCTTCCACCCCTTCCCTGCTGAGCA	25440
QY	73	CGTGAAGTAAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGT	132
DB	25439	CGTGAAGTAAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGT	25380
QY	133	GGGTGAAGTGAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGT	192
DB	25379	GGGTGAAGTGAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGT	25320
QY	193	TGACTTGCCTTAAGAAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA	252
DB	25319	TGACTTGCCTTAAGAAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA	25260
QY	253	GGAGAGCTCT-GGTGTGCTGCTGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	311
DB	25259	GGAGAGCTCTGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGA	25200
QY	312	CTTTGACTTGTGACTTTTTCACCTGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	371
DB	25199	CTTTGACTTGTGACTTTTTCACCTGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	25140
QY	372	CTGCAAGCTTCCCTGCGGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAAT	431
DB	25139	CTGCAAGCTTCCCTGCGGCTGCAATGCAATGCAATGCAATGCAATGCAATGCAAT	25080
QY	432	GGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	491
DB	25079	GGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	25021
QY	492	CGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	551
DB	25020	CGAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAG	24962
QY	552	CCCTTTCCCTGCTTGGGCTGCTGAAGAGGCTTCCCTTATCTTCCACCCCAACCT	611
DB	24961	CCCTTTCCCTGCTTGGGCTGCTGAAGAGGCTTCCCTTATCTTCCACCCCAACCT	24905
QY	612	TTCCCTCCCTGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT	671
DB	24904	TTCCCTCCCTGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCTTGGGCT	24852
QY	672	GTGAGATTCGCGGAAATTAACCTGACAGGCTTGGCAAAATGAACCTTTTGTGATCC	731
DB	24851	GTGAGATTCGCGGAAATTAACCTGACAGGCTTGGCAAAATGAACCTTTTGTGATCC	24799
QY	732	TGTACTGCTGTTTAA 748	
DB	24798	TGTACTGCTGTTTAA 24782	

RESULT 4
 AAT61405


```

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249255.
PR 17-NOV-2000; 2000US-0249257.
PR 17-NOV-2000; 2000US-0249259.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251988.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM,
PI WPI: 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 32028; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK81694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 43938 BP; 12772 A; 9656 C; 9206 G; 12304 T; 0 other;

Query Match 70.4%; Score 529.8; DB 22; Length 43938;
Best Local Similarity 91.6%; Pred. No. 9,2e-152;
Matches 675; Conservative 0; Mismatches 42; Indels 20; Gaps 10;

QY 13 TATTGAAGTAAAGAAACCTTCTCTGGAGCCTTCCACCCCTCCCTGCTGAGCA 72
DB 24420 TATTGAAGTAAAGAAACCTTCTCTGGAGCCTTCCACCCCTCCCTGCTGAGCA 24361
QY 73 CGTGAGTTCAGGAGTTCAGGAGCCTCGGAGACCTCCATGGTGGCAGAAAGGTGAGC 132
DB 24360 CGTGAGTTCAGGAGTTCAGGAGCCTCGGAGACCTCCATGGTGGCAGAAAGGTGAGC 24301
QY 133 GGGTAGTTCCTTGGCCAAAGGTGGCTCTTCAACAGAAACCAATATTTTGTTC 192

```

Pr comprising the CASP8 gene that results in inactivation of the gene
XX
PS Claim 21, Page 97-98; 107pp; English.
XX
CC Inactivation of a caspase-8 (CASP8) gene is determined by detecting a
CC modification of CASP8 genomic DNA that results in inactivation of the
CC gene. CASP8, a cysteine protease, is part of the death inducing signaling
CC complex (DISC) associated with the Fas receptor. CASP8 is inactivated in
CC cancers, and plays a role of a tumor suppressor gene. The CASP8 promoter
CC region sequences, in particular Region 1 and Region 2, are crucial to the
CC design and execution of the genomic methylation PCR analysis of CASP8
CC gene inactivation. Methylation PCR can be used to examine even minute
CC amounts of patient material to demonstrate whether the CASP8 gene
CC expresses an mRNA and protein product. The promoter Region 1 sequence is
CC located upstream (5') to exon 1, which is the alternatively-spliced 5'
CC untranslated region (UTR) that is less commonly used. The promoter
CC Region 2 sequence is located downstream (3') of exon 1 and upstream of
CC exon 2, which is the more commonly used 5' UTR. The CASP8 gene has been
CC localized to human chromosome 2q33-34. The methods are used to diagnose
CC or prognose cancer. Cancer is treated by administering a vector that
CC expresses a gene encoding functional CASP8 in cells. The cancer that
CC is diagnosed or treated is a tumour in which a myc gene is amplified,
CC such as a neuroblastoma. Aggressive neuroblastoma, juvenile neuroblastoma
CC (preferred), small-cell lung carcinoma, non-small-cell lung carcinoma,
CC colorectal carcinoma, or uterine cervical carcinoma can be diagnosed with
CC the new method. A kit for screening for a compound that induces
CC death-receptor-mediated apoptosis in cells containing an inactivated
CC CASP8 gene is also provided.
XX
SQ Sequence 753 BP; 140 A; 204 C; 171 G; 238 T; 0 other;

Query Match 100.0%; Score 753; DB 21; Length 753;
Best Local Similarity 100.0%; Pred. No. 2.3e-221;
Matches 753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTAGACCGCGTATTGAAAGTAAGAACTCTTCTCGGGAGCGCTTCCACGCCCTT 60
Db 1 AATTAGACCGCGTATTGAAAGTAAGAACTCTTCTCGGGAGCGCTTCCACGCCCTT 60

QY 61 CCCTGCTGAGACAGTGGAGTTAGGAGGAGTGGAGACTGCGATGATGTCACAG 120
Db 61 CCCTGCTGAGACAGTGGAGTTAGGAGGAGTGGAGACTGCGATGATGTCACAG 120

QY 121 AAGGGTGGAGGGGGTGGAGGCTGTGGCAAGTGCGCTTTCACAGAAACCAAT 180
Db 121 AAGGGTGGAGGGGGTGGAGGCTGTGGCAAGTGCGCTTTCACAGAAACCAAT 180

QY 181 AATTTGTTCTTGACTGCTAGAAACAGGCTGTGGGGTGGAGAACCACTTGAT 240
Db 181 AATTTGTTCTTGACTGCTAGAAACAGGCTGTGGGGTGGAGAACCACTTGAT 240

QY 241 CTGCGCTTCTGAGAGACCTCTGCTGCTGGCCAGAGTCTCTGTGTGTTCTCTC 300
Db 241 CTGCGCTTCTGAGAGACCTCTGCTGCTGGCCAGAGTCTCTGTGTGTTCTCTC 300

QY 301 TGAGCCGAGTCCCTTACTTGTACTTTTCACTCTGAGAGTCTCCAGTCTCTGCT 360
Db 301 TGAGCCGAGTCCCTTACTTGTACTTTTCACTCTGAGAGTCTCCAGTCTCTGCT 360

QY 361 ACCTTTTGTCTCCAGCTTCCCTCGCGCTCGAATGAGATACAGGAGTCCCTTCTG 420
Db 361 ACCTTTTGTCTCCAGCTTCCCTCGCGCTCGAATGAGATACAGGAGTCCCTTCTG 420

QY 421 TGAGACCGTGTGAGAGTCCAGAGACTTATCAATCCACTTTTCTTTTCTTTTG 480
Db 421 TGAGACCGTGTGAGAGTCCAGAGACTTATCAATCCACTTTTCTTTTCTTTTG 480

QY 481 GCCCTGGGGGCGACGATTAAGTACTTATCTGTCACTCTGCAATCAGCAATGCCCT 540
Db 481 GCCCTGGGGGCGACGATTAAGTACTTATCTGTCACTCTGCAATCAGCAATGCCCT 540

QY 541 GAGGTGACAGGCCCTTCCCTCTTCCGCTCTCAAGGGGTTCTTTTATGTCTTCC 600
Db 541 GAGGTGACAGGCCCTTCCCTCTTCCGCTCTCAAGGGGTTCTTTTATGTCTTCC 600

Db 541 GAGGTGACAGGCCCTTCCCTCTTCCGCTCTCAAGGGGTTCTTTTATGTCTTCC 600
QY 601 ACCCCACACCTTTCCCTCCCGCCCTGTTTGTGTGGCCCAAAACAAGTCTCTTA 660
Db 601 ACCCCACACCTTTCCCTCCCGCCCTGTTTGTGTGGCCCAAAACAAGTCTCTTA 660

QY 661 ACGTTTGGATGTGGATTGGGGAAATTAACCTGACACCGTTTGCAAAATGAACCTTTT 720
Db 661 ACGTTTGGATGTGGATTGGGGAAATTAACCTGACACCGTTTGCAAAATGAACCTTTT 720

QY 721 TTTTGTGATCCTGTACACTGTTTTTTACCTT 753
Db 721 TTTTGTGATCCTGTACACTGTTTTTTACCTT 753

RESULT 2
AAK77216/C
ID AAK77216 standard; DNA; 43938 BP.
XX
XX AAK77216;
AC
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32028.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
PD
XX
XX 09-AUG-2001.
PF
XX 17-JAN-2001; 2001WO-US01354.
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 751.255 Seconds
(without alignments)
2257.229 Million cell updates/sec

Title: US-09-477-082-2

Perfect score: 753
Sequence: 1 aatagacgcgtatgaaa.....tacactgttttttaacctt 753

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	753	100.0	753	21	AA51793
2	529.8	70.4	43938	22	AAK77216
3	529.8	70.4	45017	22	AAK77217
4	123.6	16.4	2887	18	AA61405
5	123.6	16.4	2887	12	AA03915
6	123.6	16.4	2887	24	ABK84723
7	60	8.0	60	24	ABN58902
8	39.4	5.2	9741	24	ABL33323
9	38.4	5.1	8300	22	AAH25711

10	38	5.0	439	22	AA193505	Human polynucleoti
C 11	37.4	5.0	1227	21	AAC36552	Arabidopsis thalia
C 12	37.4	5.0	1522	21	AAC51060	Arabidopsis thalia
C 13	37.2	4.9	4983	23	ABL03779	Drosophila melanog
C 14	37.2	4.9	7723	23	ABL03778	Drosophila melanog
C 15	36.4	4.8	10732	21	AA10594	Gene encoding a su
C 16	35.8	4.8	700	22	AAH92818	Human inflammatory
C 17	35.8	4.8	700	22	AAH92819	Human inflammatory
C 18	35.8	4.8	9743	22	AA104517	Human reproductive
C 19	35.8	4.8	90220	24	ABR83576	Human CDNA differe
C 20	35.4	4.7	2291	23	ABL10669	Drosophila melanog
C 21	35.4	4.7	3079	23	ABL08820	Drosophila melanog
C 22	35.4	4.7	4453	23	ABL10668	Drosophila melanog
C 23	35.2	4.6	2603	24	ABR65157	Arabidopsis CDNA e
C 24	35	4.6	6146	22	ABA07752	Human ovarian and
C 25	35	4.6	744	22	AA103541	Human reproductive
C 26	34.8	4.6	6146	22	AA195563	Human neuroblastom
C 27	34.8	4.6	6741	21	AA10585	Gene encoding a su
C 28	34.8	4.6	30310	24	AA18253	Genomic DNA from t
C 29	34.6	4.6	6277	21	AA18253	Human OREF ORF2414
C 30	34.4	4.6	448	22	AA180552	Human polynucleoti
C 31	34.4	4.6	2779	24	ABR84714	Human CDNA differe
C 32	34.4	4.6	3059	22	AA158335	Human polynucleoti
C 33	34.4	4.6	3068	22	AA158336	Human polynucleoti
C 34	34.4	4.6	3071	22	AA158337	Human polynucleoti
C 35	34.4	4.6	3134	22	AA158334	Human polynucleoti
C 36	34.4	4.6	3713	21	AA18160	Human cancer assoc
C 37	34.4	4.6	126512	24	ABN83429	Human transporter
C 38	34.2	4.5	1494	24	ABR63596	Rat sequence diffe
C 39	34	4.5	143068	21	AA121105	Human low adenosin
C 40	34	4.5	143068	21	AA121105	Human low adenosin
C 41	34	4.5	143068	21	AA121105	Human low adenosin
C 42	34	4.5	143068	21	AA121105	Human low adenosin
C 43	34	4.5	143068	24	AA158124	Human adenosine re
C 44	34	4.5	149412	21	AA135151	Ovary cancer relat
C 45	34	4.5	132740	21	AA121273	Human adenosine re

ALIGNMENTS

AA51793	standard; DNA; 753 BP.
AA51793	
AA51793	
31-OCT-2000	(first entry)
CASP8 promoter Region 2.	
CASP8; caspase-8; promoter; Region 2; inactivation; methylation;	
Cysteine protease; tumour suppressor; chromosome 2q33-34; neuroblastoma;	
cancer; death receptor; apoptosis; cytosolic; gene therapy; ss.	
XX	
OS	Homo sapiens.
XX	
FN	WO200039347-A1.
XX	
PD	06-JUL-2000.
XX	
PF	30-DEC-1999; 99WO-051280.
XX	
PR	31-DEC-1998; 98US-0114308.
XX	
PA	(STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX	
PI	Kidd VJ, Lantti JM, Teltz T;
XX	
DR	WPI: 2000-452423/39.
XX	
PT	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
	prognosing cancer, comprises detecting a modification of genomic DNA

[illegible]

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
Williams, G., Williamson, A., Wlarczyk, R., Wooden, S., Worley, K.,
Wu, C., Wu, X., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.

Unpublished
2 (bases 1 to 128941)

Worley, K.C.

Direct Submission
Submitted (11-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 128941)

Worley, K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18847091.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GRX

Center clone name: CH230-257E5

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 84896 bases at least Q40

Consensus quality: 89146 bases at least Q30

Consensus quality: 92809 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 55 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1422: contig of 1422 bp in length
1423 1522: gap of unknown length
1523 2531: contig of 1009 bp in length
2532 2631: gap of unknown length
2631 3933: contig of 1302 bp in length
3933 4033: gap of unknown length
4033 5147: contig of 1114 bp in length
5147 5247: gap of unknown length
5247 5375: contig of 1128 bp in length
5375 6475: gap of unknown length
6475 7682: contig of 1207 bp in length
7682 7783: gap of unknown length
7783 8958: contig of 1176 bp in length
8958 9058: gap of unknown length
9058 10124: contig of 1066 bp in length
10124 10224: gap of unknown length
10224 11431: contig of 1207 bp in length
11431 11531: gap of unknown length
11531 12623: contig of 1092 bp in length
12623 12723: gap of unknown length
12723 14127: contig of 1404 bp in length
14127 14227: gap of unknown length
14227 15363: contig of 1136 bp in length
15363 15463: gap of unknown length
15463 16626: contig of 1163 bp in length

16627 16726: gap of unknown length
16727 17904: contig of 1178 bp in length
17904 18004: gap of unknown length
18004 19198: contig of 1194 bp in length
19198 19298: gap of unknown length
19298 20450: contig of 1152 bp in length
20450 20551: gap of unknown length
20551 21588: contig of 1038 bp in length
21588 21689: gap of unknown length
21689 22697: contig of 1009 bp in length
22697 22797: gap of unknown length
22797 23963: contig of 1166 bp in length
23963 24063: gap of unknown length
24063 26414: contig of 2351 bp in length
26414 26514: gap of unknown length
26514 27791: contig of 1277 bp in length
27791 27891: gap of unknown length
27891 29601: contig of 1710 bp in length
29601 29701: gap of unknown length
29701 31394: contig of 1693 bp in length
31394 31494: gap of unknown length
31494 32917: contig of 1423 bp in length
32917 33017: gap of unknown length
33017 34819: contig of 1802 bp in length
34819 34919: gap of unknown length
34919 37158: contig of 2239 bp in length
37158 37258: gap of unknown length
37258 39046: contig of 1788 bp in length
39046 39146: gap of unknown length
39146 41095: contig of 1949 bp in length
41095 41196: gap of unknown length
41196 41965: contig of 2270 bp in length
41965 43465: gap of unknown length
43465 44580: contig of 1015 bp in length
44580 44681: gap of unknown length
44681 47162: contig of 2482 bp in length
47162 47262: gap of unknown length
47262 48007: contig of 1645 bp in length
48007 49007: gap of unknown length
49007 50894: contig of 1887 bp in length
50894 50994: gap of unknown length
50994 52932: contig of 1938 bp in length
52932 53032: gap of unknown length
53032 54663: contig of 1631 bp in length
54663 54763: gap of unknown length
54763 56965: contig of 2202 bp in length
56965 57065: gap of unknown length
57065 59664: contig of 2699 bp in length
59664 59864: gap of unknown length
59864 62414: contig of 2550 bp in length
62414 62514: gap of unknown length
62514 65241: contig of 2727 bp in length
65241 65341: gap of unknown length
65341 67570: contig of 2229 bp in length
67570 67670: gap of unknown length
67670 70493: contig of 2823 bp in length
70493 70593: gap of unknown length
70593 72869: contig of 2376 bp in length
72869 73069: gap of unknown length
73069 75496: contig of 2427 bp in length
75496 75596: gap of unknown length
75596 77828: contig of 2232 bp in length
77828 77928: gap of unknown length
77928 81696: contig of 3768 bp in length
81696 81796: gap of unknown length
81796 84238: contig of 2442 bp in length
84238 84338: gap of unknown length
84338 86916: contig of 2578 bp in length
86916 87016: gap of unknown length
87016 90937: contig of 3921 bp in length
90937 91037: gap of unknown length
91037 95061: contig of 4024 bp in length
95061 95161: gap of unknown length

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="peripheral blood lymphocyte"
86. .916

```

Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, L.,

KEYWORDS	SOURCE	ORGANISM	cell death; MACH-alpha-1; protease.
REFERENCE	AUTHORS	Titile	Homo sapiens.
REFERENCE	AUTHORS	Titile	Homo sapiens.
REFERENCE	AUTHORS	Titile	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	AUTHORS	Titile	1 (bases 1 to 2887)
REFERENCE	AUTHORS	Titile	Boldin, M. P., Goncharov, T. M., Golse, V. V. and Wallach, D.
REFERENCE	AUTHORS	Titile	Involvement of MACH, a novel MOR1/PAD2-interacting protease, in Fas/APO-1- and TNF receptor-induced cell death
REFERENCE	AUTHORS	Titile	1 (bases 1 to 2887)
REFERENCE	AUTHORS	Titile	8681376
REFERENCE	AUTHORS	Titile	2 (bases 1 to 2887)
REFERENCE	AUTHORS	Titile	Wallach, D.
REFERENCE	AUTHORS	Titile	Direct Submission
REFERENCE	AUTHORS	Titile	Submitted (24-MAY-1996) D. Wallach, The Weizmann Institute, Dept of
REFERENCE	AUTHORS	Titile	Membrane Research & Biophysics, Rehovot 76100, ISRAEL
REFERENCE	AUTHORS	Titile	Revised by [3]
REFERENCE	AUTHORS	Titile	3 (bases 1 to 2887)
REFERENCE	AUTHORS	Titile	Wallach, D.
REFERENCE	AUTHORS	Titile	Direct Submission
REFERENCE	AUTHORS	Titile	Submitted (17-JUN-1996) D. Wallach, The Weizmann Institute, Dept of
REFERENCE	AUTHORS	Titile	Membrane Research & Biophysics, Rehovot 76100, ISRAEL
REFERENCE	AUTHORS	Titile	4 (bases 1 to 2887)
REFERENCE	AUTHORS	Titile	Mandruzzato, S., Brasseur, F., Andry, G., Boon, T. and van der
REFERENCE	AUTHORS	Titile	Bruggen, P.
REFERENCE	AUTHORS	Titile	A CASP-8 mutation recognized by cytolytic T lymphocytes on a human
REFERENCE	AUTHORS	Titile	head and neck carcinoma
REFERENCE	AUTHORS	Titile	J Exp. Med. 186 (5), 785-793 (1997)
REFERENCE	AUTHORS	Titile	97419196
REFERENCE	AUTHORS	Titile	9271594
REFERENCE	AUTHORS	Titile	Location/Qualifiers
REFERENCE	AUTHORS	Titile	1. .2887
REFERENCE	AUTHORS	Titile	/organism="Homo sapiens"
REFERENCE	AUTHORS	Titile	/db_xref="taxon:9606"
REFERENCE	AUTHORS	Titile	/tissue_type="thymus"
REFERENCE	AUTHORS	Titile	292. .1731
REFERENCE	AUTHORS	Titile	/function="protease"
REFERENCE	AUTHORS	Titile	/note="triggers cell death"
REFERENCE	AUTHORS	Titile	/codon_start=1
REFERENCE	AUTHORS	Titile	/product="MACH-alpha-1"
REFERENCE	AUTHORS	Titile	/protein_id="CA66853.1"
REFERENCE	AUTHORS	Titile	/db_xref="GI:2440071"
REFERENCE	AUTHORS	Titile	/db_xref="SWISS-PROT:Q14790"
REFERENCE	AUTHORS	Titile	/translation="MPSRLYDGEOLDESLASLFLSLDYIPKROEPKIDALMIL
REFERENCE	AUTHORS	Titile	FORLOEKRMLESNLSFLKELPRINLDLITLINTKREMERLOTGPAQISARY
REFERENCE	AUTHORS	Titile	VMYQISEEVSSEIRSEFLLOEIEISKCLDDMDMLDIIFEMEKVITIGEGKDLIT
REFERENCE	AUTHORS	Titile	KRYCAQINRSKLITINDYEEFSKRSSELSGSPRENGEELCVMTISSPRODSE
REFERENCE	AUTHORS	Titile	SQILDYQKRSKPRGICILINNEFKAREKPKLHSIRDRNTHLDAGLITTFEE
REFERENCE	AUTHORS	Titile	LHEIKRHDCTVEQIETELIKIYOLMNSMNDCTCCILSHGDKGIYGVDSQAPARY
REFERENCE	AUTHORS	Titile	ELTSPGFLGKPSLAGKPKVEFLIOACGQYOKIKIPETSDSEQPYEMDLSPTQY
REFERENCE	AUTHORS	Titile	IPEADPLGLMATVNNVCVSRNPAGEWYIOSLQGSRLRECRPGDDITLTLEVEV
REFERENCE	AUTHORS	Titile	SNKDKNMGMKOMPFTLRKLVKPFSD"
REFERENCE	AUTHORS	Titile	BASE COUNT
REFERENCE	AUTHORS	Titile	653 a 595 c 625 g 814 t
REFERENCE	AUTHORS	Titile	ORIGIN
REFERENCE	AUTHORS	Titile	Query Match
REFERENCE	AUTHORS	Titile	Best Local Similarity 93.5%; Pred. No. 4.3e-27;
REFERENCE	AUTHORS	Titile	Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
REFERENCE	AUTHORS	Titile	13 TATGAAGTAAAGAAAGAACTTCTCCGCGGAGCCTTCCACCCCTTCTGCTGAGCA 72
REFERENCE	AUTHORS	Titile	144 TATTGAAAGTAGAAGAAACTTCTTCGCGGAGCCTTCCACCCCTTCTGCTGAGCA 203
REFERENCE	AUTHORS	Titile	73 CGTGGAGTTAGGAGAGCTTAGGGAGCTGGAGACTGCGATGGTGCAGAGAAAGCTGAGC 132
REFERENCE	AUTHORS	Titile	204 CGTGGAGTTAGGAGAGCTTAGGGAGCTGGAGAGCTGCGATGGTGCAGAGAAAGCTGAGC 263
REFERENCE	AUTHORS	Titile	133 GGTGGAGTGCCTGTTCGC 150
REFERENCE	AUTHORS	Titile	264 GGATTATATTTCTCTCTGCC 281

AF422929		777 bp	mRNA	linear	PRI 03-Apr-2002
LOCUS	Homo sapiens clone 4 procaspase-8 (CASP8)		mRNA, partial cds:		
DEFINITION	alternatively spliced.				
ACCESSION	AF422929				
VERSION	AF422929.1 GI:19401529				
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens.				
	Homo sapiens.				
	Eumalvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Euthelia; Primates; Carnivora; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 777)				
AUTHORS	Breckenridge,D.G., Nguyen,M., Kuppiq,S., Reth,M. and Shore,G.C.				
TITLE	The procaspase-8 isoform, procaspase-8L, recruited to the BAP31 complex at the endoplasmic reticulum				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)				
MEDLINE	21927603				
PUBMED	11917123				
REFERENCE	2 (bases 1 to 777)				
AUTHORS	Breckenridge,D.G., Nguyen,M., Kuppiq,S., Reth,M. and Shore,G.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-SEP-2001) Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada				
FEATURES	Location/Qualifiers				
source	1..777				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="4"				
gene	1..>777				
	/gene="CASP8"				
CDS	266..>777				
	/gene="CASP8"				
	/note="caspase precursor; alternatively spliced"				
	/codon_start=1				
	/product="procaspase-8"				
	/protein_id="AAL87632.1"				
	/db_xref="GI:19401530"				
	/translation="MDFSRNLYDYDGEDLDSEDLAKFLSDYIPQRKPEIKDALMLTILFQROEKRLSESNLSFKLELFRLINLDLITYLTNRKEEMSELDTGPAAOISAYRVMLOIIEEVSRSLSRFKFLIQEIRISKDDMDNMNLDFIERKRVIIIGEGKLILLKRAAOIKSL"				
BASE COUNT	218 a 156 c 209 g 194 t				
ORIGIN					
	Query Match	16.2%; Score 121.8; DB 9; Length 777;			
	Best Local Similarity 98.4%; Pred. No. 1.4e-26;				
Matches	123; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	13 TATTGAAGTTAAGAAGAACTTCCTTCCTGGGAGCCCTTCCACCCTTCCTCGTGAGCA 72				
Db	144 TATTGAAGTTAAGAAGAACTTCCTTCCTGGGAGCCCTTCCACCCTTCCTCGTGAGCA 203				
QY	73 CGTAGACTTAGGCAGCTTAGGGAGCTGGAGACTGTGCATGTGCCAGAAAGGTGGAGC 132				
Db	204 CGTAGACTTAGGCAGCTTAGGGAGCTGGAGACTGTGCATGTGCCAGAAAGGTGGAGC 263				
QY	133 GGCTG 137				
Db	264 GGATG 268				
RESULT 10					
LOCUS	166494	7218 bp	DNA	linear	PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.				
ACCESSION	166494				
VERSION	166494.1 GI:2724471				
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
	1 (bases 1 to 7218)				

REFERENCE 2 (bases 1 to 2503)
 Breckenridge,D.G., Nguyen,M., Kupzig,S., Reth,M. and Shore,G.C.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-2001) Biochemistry, McGill University, 3655
 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada
 FEATURES
 SOURCE 1..2503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 gene 1..2503
 /gene="CASP8"
 115..1731
 /gene="CASP8"
 /note="caspase precursor; alternatively spliced"
 /codon_start=1
 /product="procaspase-8"
 /protein_id="AA18728.1"
 /db_xref="GI:19401519"
 /translation="MEGRRARVIESRNFELGAPPPPAHVHVLGRIGDSEETAM
 PKRGADYLLPFRKMDPSRLNYDIGEOLDSGLSLKSLDYIPORKEPIKDALM
 LFORLOEKRLSESLFLKELLFRINRLDITYINTRKEEMERELQPGRAOISAV
 RVLVYISEVSRSELSRFLIOEIRISCKLDDMNLDITEMKRVIIIGEGKIDI
 LKRVCAQINKSLIKITINDYEBERSKSSLSGSPDFSGEELCGVMTISDSPREDS
 ESQTLKRVYQMSKPRGYCLIIINHNFAKAREVPLHSIRDRNGTHLDAGALTTFE
 ELHEIKRPHDDCTVEQIYEILKTYQLMDSNMDCTICLLSGDKGIIVGTGQAPVI
 YELTSQFTGLKCPSLAGKPRVFIOACQDNYOKGIPVETDSEEQYLEMDLSSQTR
 YIPDEADFLGMATVNNVCVSYRNPAEGTWYIOSLCSLRERCPRGDDILITILEVNYE
 VSNKDKKMNQKMPQPTFLRKXIVFSPD"

BASE COUNT 726 a 528 c 593 g 656 t
 ORIGIN

Query Match 16.4%; Score 123.6; DB 9; Length 2503;
 Best Local Similarity 93.5%; Pred. No. 4.2e-27;
 Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCGGAGACCTTCCACCCCTTCCTGCTAGACA 72
 |||||||
 DB 144 TATTGAAGTAAAGAACTTCTCTCGGAGACCTTCCACCCCTTCCTGCTAGACA 203
 |||||||

QY 73 CGTGAAGTTAGGAGGACTCGAGACTGCGATGTCGACAGAAAGGATGAGAC 132
 |||||||
 DB 204 CGTGAAGTTAGGAGGACTCGAGACTGCGATGTCGACAGAAAGGATGAGAC 263
 |||||||

QY 133 GGGTGAAGTCCCTGTGCC 150
 |||||
 DB 264 GGATTATATCTCTCTGCC 281
 |||||

RESULT 6
 AR211527
 LOCUS AR211527 2887 bp DNA linear PAT 20-JUN-2002
 DEFINITION Sequence 14 from patent US 6399327.
 ACCESSION AR211527
 VERSION AR211527.1 GI:21514872
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2887)
 AUTHORS Wallach,D., Boidin,M., Goncharov,T. and Golstrev,Y.V.
 TITLE Modulators of the function of FAS receptors and other proteins
 JOURNAL Patent: US 6399327-A 14 04-JUN-2002;
 FEATURES
 SOURCE 1..2887
 /organism="unknown"
 BASE COUNT 853 a 595 c 625 g 814 t
 ORIGIN

Query Match 16.4%; Score 123.6; DB 6; Length 2887;
 Best Local Similarity 93.5%; Pred. No. 4.3e-27;
 Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCGGAGACCTTCCACCCCTTCCTGCTAGACA 72

DB 144 TATTGAAGTAAAGAACTTCTCTCGGAGACCTTCCACCCCTTCCTGCTAGACA 203
 |||||||

QY 73 CGTGAAGTTAGGAGGACTCGGAGACTGCGATGTCGACAGAAAGGATGAGAC 132
 |||||||

DB 204 CGTGAAGTTAGGAGGACTCGGAGACTGCGATGTCGACAGAAAGGATGAGAC 263
 |||||||

QY 133 GGGTGAAGTCCCTGTGCC 150
 |||||

DB 264 GGATTATATCTCTCTGCC 281
 |||||

RESULT 7
 AX134419
 LOCUS AX134419 2887 bp DNA linear PAT 29-MAY-2001
 DEFINITION Sequence 15 from Patent WO0129232.
 ACCESSION AX134419
 VERSION AX134419.1 GI:14271065
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2887)
 AUTHORS Cordell,B. and Li,Y.
 TITLE Functional cloning of genes encoding proteins/enzymes involved in
 proteolytic cleavage
 JOURNAL Patent: WO 0129232-A 15 26-Apr-2001;
 Scios Inc. (US)
 FEATURES
 SOURCE Location/Qualifiers
 1..2887
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 292..1731
 CDS
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC39526.1"
 /db_xref="GI:14271066"
 /translation="MDFSRNLYDIGEOLDSGLSLKSLDYIPORKEPIKDALM
 FOLQEKRLSESLFLKELLFRINRLDITYINTRKEEMERELQPGRAOISAV
 RVLVYISEVSRSELSRFLIOEIRISCKLDDMNLDITEMKRVIIIGEGKLDIL
 KRVCAQINKSLIKITINDYEBERSKSSLSGSPDFSGEELCGVMTISDSPREDS
 ESQTLKRVYQMSKPRGYCLIIINHNFAKAREVPLHSIRDRNGTHLDAGALTTFE
 ELHEIKRPHDDCTVEQIYEILKTYQLMDSNMDCTICLLSGDKGIIVGTGQAPVI
 YELTSQFTGLKCPSLAGKPRVFIOACQDNYOKGIPVETDSEEQYLEMDLSSQTR
 YIPDEADFLGMATVNNVCVSYRNPAEGTWYIOSLCSLRERCPRGDDILITILEVNYE
 VSNKDKKMNQKMPQPTFLRKXIVFSPD"

BASE COUNT 853 a 595 c 625 g 814 t
 ORIGIN

Query Match 16.4%; Score 123.6; DB 6; Length 2887;
 Best Local Similarity 93.5%; Pred. No. 4.3e-27;
 Matches 129; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 13 TATTGAAGTAAAGAACTTCTCTCGGAGACCTTCCACCCCTTCCTGCTAGACA 72
 |||||||

DB 144 TATTGAAGTAAAGAACTTCTCTCGGAGACCTTCCACCCCTTCCTGCTAGACA 203
 |||||||

QY 73 CGTGAAGTTAGGAGGACTCGGAGACTGCGATGTCGACAGAAAGGATGAGAC 132
 |||||||

DB 204 CGTGAAGTTAGGAGGACTCGGAGACTGCGATGTCGACAGAAAGGATGAGAC 263
 |||||||

QY 133 GGGTGAAGTCCCTGTGCC 150
 |||||

DB 264 GGATTATATCTCTCTGCC 281
 |||||

RESULT 8
 HSMACHA1
 LOCUS HSMACHA1 2887 bp mRNA linear PRI 25-SEP-1997
 DEFINITION H.sapiens mRNA for MACH-alpha-1 protein.
 ACCESSION X98172
 VERSION X98172.1 GI:1403318

QY 73 CGTGAAGTTAGGAGGAGTGGGAGCTCGAGACTCGATGTCGCCAGAAAGGTGAGC 132
 DB 1682 CGTGAAGTTAGGAGGAGTGGGAGCTCGAGACTCGATGTCGCCAGAAAGGTGAGC 1741
 QY 133 GGGTAGTGTCCCTGTGGCCAGGTGCTCTTAAAGAGAAACCAATATTTTCTTCT 192
 DB 1742 GGGTAGTGTCCCTGTGGCCAGGTGCTCTTAAAGAGAAACCAATATTTTCTTCT 1801
 QY 193 TGAATGCTCTGAAGAAACAGGCTGTGGGGGTGGGAGCACTTGATGTCCTTGA 252
 DB 1802 TGACTTGCTCTGAAGAAACAGGCTGTGGGGGTGGGAGCACTTGATGTCCTTGA 1861
 QY 253 GGAACCTCTT-GGTGCTGCTGGCCAGGTGCTCTGTGTGTTCTGTGAGCCGATG 311
 DB 1862 GGAACCTCTGCTGTGCTGCTGGCCAGGTGCTCTGTGTGTTCTGTGAGCCGATG 1921
 QY 312 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371
 DB 1922 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1981
 QY 372 CTCGAAGCTTCCCTGCGCCCTGGAATGAGATACAGGACTCCCTTGTGAGCCGAT 431
 DB 1982 CTCGAAGCTTCCCTGCGCCCTGGAATGAGATACAGGACTCCCTTGTGAGCCGAT 2041
 QY 432 GGAGAGTCCAGAGACTTATATCAATCCACTTTTCTTTTCAATTTGGCCCTGGG 491
 DB 2042 GGAGAGTCCAGAGACTTATATCAATCCACTTTTCTTTTCAATTTGGCCCT 2100
 QY 492 CGACGGTTAGTACTTATCTGTCATCTGTCGATACAGAAAGCCCGAGGTGAGAG 551
 DB 2101 CGACGGTTAGTACTTATCTGTCATCTGTCGATACAGAAAGCCCGAGGTGAGAG 2151
 QY 552 CCCCTTCCCTCTTCTTCCGCTGCTGAGAGGCTTCTTCTTCTTCTTCTTCTTCT 611
 DB 2152 CCCCTTCCCTCTTCTTCCGCTGCTGAGAGGCTTCTTCTTCTTCTTCTTCTTCT 2208
 QY 612 TTCCCTTCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 671
 DB 2209 TTCCCTTCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2261
 QY 672 GTGATTCGCGGAAATTAACCTGCACCCGTTTGAATGAATGACTTTTCTTCTTCT 731
 DB 2262 GTGATTCGCGGAAATTAACCTGCACCCGTTTGAATGAATGACTTTTCTTCTTCT 2314
 QY 732 TGTACACTGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 748
 DB 2315 TGTACACTGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2331
 RESULT 4
 AF210257 962 bp DNA linear PRI 06-JUN-2000
 LOCUS Homo sapiens caspase 8 gene, 5' flanking region.
 DEFINITION AF210257
 ACCESSION AF210257.1 GI:8272474
 VERSION
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE
 AUTHORS Teltz,T., Mel,T., Valentine,M.B., Vanin,E.F., Grenet,J.,
 Caspase 8 is deleted or silenced preferentially in childhood
 neuroblastomas with amplification of MYCN
 Nat. Med. 6 (5), 529-535 (2000)
 JOURNAL MEDLINE
 PUBMED 10802708
 REFERENCE 2 (bases 1 to 982)
 AUTHORS Teltz,T., Mel,T., Valentine,M.B., Vanin,E.F., Grenet,J.,
 Valentine,V.A., Behm,F.G., Look,A.T., Lantti,J.M. and Kidd,V.J.
 TITLE Direct Submission

JOURNAL Submitted (01-DEC-1999) Tumor Cell Biology, St. Jude Children's
 Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA
 FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="q33-q34"
 misc_feature 1..982
 /note="5' flanking region of caspase 8 gene"
 BASE COUNT 198 a 235 c 241 g 308 t
 ORIGIN
 Query Match 64.9%; Score 488.4; DB 9; Length 982;
 Best Local Similarity 98.5%; Pred. No. 2,1e-141;
 Matches 514; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
 QY 13 TATTGAAGTAAAGAAACTTCTTCTGAGAGCTTTCCACCCCTTCCCTGCTGAGCA 72
 DB 459 TATTGAAGTAAAGAAACTTCTTCTGAGAGCTTTCCACCCCTTCCCTGCTGAGCA 518
 QY 73 CGTGAAGTTAGGAGGAGTGGGAGCTGAGAGCTGCGATGCTGCCAGAAAGGTGAGC 132
 DB 519 CGTGAAGTTAGGAGGAGTGGGAGCTGAGAGCTGCGATGCTGCCAGAAAGGTGAGC 578
 QY 133 GGGTAGTGTCCCTGTGGCCAGGTGCTCTTCAACAGAGAAACCAATATTTTGTTC 192
 DB 579 GGGTAGTGTCCCTGTGGCCAGGTGCTCTTCAACAGAGAAACCAATATTTTGTTC 638
 QY 193 TGACTTGCTCTGAGAAACAGGCTGTGGGGGTGGGAGCACTTGATCTGCCCTTCTGA 252
 DB 639 TGACTTGCTCTGAGAAACAGGCTGTGGGGGTGGGAGCACTTGATCTGCCCTTCTGA 698
 QY 253 GGAACCTCTT-GGTGCTGCTGGCCAGGTGCTCTGCTGCTGCTGCTGCTGCTGCTG 311
 DB 699 GGAACCTCTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 758
 QY 312 CTGTGACTTCTCTTCTTCTTCTGCTGAGAGCTCCAGTCTCTGCTGCTGCTGCTGCT 371
 DB 759 CTGTGACTTCTCTTCTTCTTCTGCTGAGAGCTCCAGTCTCTGCTGCTGCTGCTGCT 818
 QY 372 CTCGAAGTTCCCTGCGCCCTGGAATGAGATACAGGACTCCCTTGTGGAGCCGATG 431
 DB 819 CTCGAAGTTCCCTGCGCCCTGGAATGAGATACAGGACTCCCTTGTGGAGCCGATG 878
 QY 432 GGAAGTCCAGAGACTTATATCAATCCACTTTTCTTCTTCTTCTTCTTCTTCTTCT 491
 DB 879 GGAAGTCCAGAGACTTATATCAATCCACTTTTCTTCTTCTTCTTCTTCTTCTTCT 937
 QY 492 CGACGGTTAAGTACTTATCTGTCATCTGTCGATGCTGCAATCAGCA 533
 DB 938 CGACGGTTAAGTACTTATCTGTCATCTGTCGATGCTGCAATCAGCA 979
 RESULT 5
 AF422925 2503 bp mRNA linear PRI 03-APR-2002
 LOCUS Homo sapiens procaspase-8L (CASP8) mRNA, complete cds;
 DEFINITION AF422925
 ACCESSION AF422925
 VERSION
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 TITLE
 AUTHORS Breckenridge,D.G., Nguyen,M., Knapp,S., Beth,M. and Shore,G.C.
 The procaspase-8 isoform, procaspase-8L, recruited to the BAP1
 complex at the endoplasmic reticulum
 Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)
 JOURNAL MEDLINE
 PUBMED 11917123

```

misc-feature      5194..5761
                  /note="similar to Homo sapiens EST BF529656
                  (NID:g1617019)"
misc-feature      5237..6019
                  /note="similar to Homo sapiens EST AU130508
                  (NID:g10990862)"
misc-feature      5238..6043
                  /note="similar to Mus musculus EST BF166846
                  (NID:g1047198)"
misc-feature      5326..5653
                  /note="similar to Homo sapiens EST BG390376
                  (NID:g13283624)"
misc-feature      5361..6051
                  /note="similar to Homo sapiens EST AU135897
                  (NID:g10996436)"
misc-feature      5361..5812
                  /note="similar to EST BG662638 (NID:g14213176)"
misc-feature      5420..6048
                  /note="similar to Homo sapiens EST BE788113
                  (NID:g10209311)"
misc-feature      5551..6327
                  /note="similar to Homo sapiens EST AU130771
                  (NID:g10991125)"
misc-feature      5615..55984
                  /note="similar to Homo sapiens EST AA280659 (NID:g1923454)

Query Match      70.4%; Score 529.8; DB 9; Length 181150;
Best Local Similarity 91.6%; Pred. 3.9e-154;
Matches 675; Conservative 0; Mismatches 42; Indels 20; Gaps 10;

QY      13 TATTGAAGTAAAGAACTTCTTCTGAGACCTTCCACCCCTTCCCTGCTGAGCA 72
DB      159209 TATTGAAGTAAAGAACTTCTTCTGAGACCTTCCACCCCTTCCCTGAGCA 159150

QY      73 CGTGGAGTAGGAGAGGTGGGAGCTCGAGACTCGAGATGATGATGCGCAGAAAGGTGAGAC 132
DB      159149 CGTGGAGTAGGAGAGGTGGGAGCTCGAGACTCGAGATGATGATGCGCAGAAAGGTGAGAC 159090

QY      133 GGGTAGTGCCCTGTTGCCAAGGTGGCTTTTCAACAGAAACCAATATTTTGTTCCT 192
DB      159089 GGGTAGTGCCCTGTTGCCAAGGTGGCTTTTCAACAGAAACCAATATTTTGTTCCT 159030

QY      193 TGACTGCTCTAGAAACAGAGGCTGTGGGGGTGGGAGAACTTGGATGCTGCCCTTCA 252
DB      159029 TGACTGCTCTAGAAACAGAGGCTGTGGGGGTGGGAGAACTTGGATGCTGCCCTTCA 158970

QY      253 GGACACCTCT-GGTGTGCTGTGGCCAGGTCTCTGTGTGTTTCTCTGAGCGATGC 311
DB      158969 GGACACCTCTGTGGGTGCTGTGGCCAGGTCTCTGTGTGTTTCTCTGAGCGATGC 158910

QY      312 CTTTGACTTGTCTACTTTTCACTTGACAGCTCTCCACTTCTCTGCTACTCTTTTGTG 371
DB      158909 CTCTGACTTGTCTACTTTTCACTTGACAGCTCTCCACTTCTCTGCTACTCTTTTGTG 158850

QY      372 CTCGAGCTTCCCTGCGCCCTCGAATCGATACAGAGCTCCCTCTGTGACCCGTT 431
DB      158849 CTCGAGCTTCCCTGCGCCCTCGAATCGATACAGAGCTCCCTCTGTGACCCGTT 158790

QY      432 GGAAGATCCAGAAAGACTTATCATCCACTTTTCTTTTCAATTGAGCCCTGGGCGC 491
DB      158789 GGAAGATCCAGAAAGACTTATCATCCACTTTTCTTTTCAATTGAGCCCT-GGGCGC 158731

QY      492 CGAGGGTAGTAGTCTTATCTGTCTGTCATCTGTGCGAATCGAATGCGCCGTGAGTGCAGC 551
DB      158730 CGAGGGTAGTAGTCTTATCTGTCTGTCATCTGTGCGAATCGAATGCGCCGTGAGTGCAGC 158672

QY      552 CCCCTTCCCTCTTTCGGGCTCCTGAAGGGGTTTCCTTTATGTCTTCCACCCCAACCT 611
DB      158671 CCCCTTCCCTCTTTCGGGCTCCTGAAGGGGTTTCCTTTATGTCTTCCACCCCAACCT 158615

QY      612 TTTCCCTCCCTGCTCTGTTTGTGTGCCCCAAGATGTTCTTAACGTTTTCAT 671
DB      158614 TTTCCCTCCCTGCTCTGTTTGTGTGCCCCAAGATGTTCTTAACGTTTTCAT 158562

```

```

QY      672 GTGATTCGCGAAATTAATTAACCTGCACCCGTTTCGAAATGAACTTTTGTGATCC 731
DB      158561 GTGATTCGTC---AATTATCTGTATC--TTGCAAAATGTA--TTTTCCTTTGATG 158509

QY      732 TGTACACTGTTTTTTTA 748
DB      158508 TGTACACTGTTTTTTTA 158492

RESULT 3
AB03898052
LOCUS      AB03898052 6502 bp DNA linear PRI 23-FEB-2001
DEFINITION Homo sapiens gene for caspase-8, exon 3, exon 4.
ACCESSION AB038981
VERSION    AB038981.1 GI:12862688
KEYWORDS
SEGMENT
SOURCE
ORGANISM   Homo sapiens DNA.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS    Hadano,S., Yanagisawa,Y., Skaug,J., Fichter,K., Nasir,J.,
            Martindale,D., Koop,B.E., Scherer,S.W., Nicholson,D.W.,
            Rouleau,G.A., Ikeda,J.-E. and Hayden,M.R.
TITLE      Cloning and characterization of three novel genes, ALS2CR1,
            ALS2CR2, and ALS2CR3, in the juvenile amyotrophic lateral sclerosis
            (ALS2) critical region at chromosome 2q33-q34: candidate genes for
            ALS2
JOURNAL    Genomics 71 (2), 200-213 (2001)
MEDLINE    21100893
REFERENCE   2 (bases 1 to 6502)
AUTHORS    Hadano,S., Ikeda,J. and Hayden,M.R.
TITLE      Direct Submission
JOURNAL    Submitted (24-FEB-2000) Shinji Hadano, Tokai University, The
            Institute of Medical Sciences, Bohseidai, Isehara, Kanagawa
            259-1193, Japan (E-mail:shinji.engi.med.u-tokai.ac.jp,
            Tel:+81-463-91-5095, Fax:+81-463-91-4993)
FEATURES
source
    1..6502
    /organism="Homo Sapiens"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2q33"
    <1..1478
    /note="intron is approximately 24.0 kbp long"
    /number=2
    1479..1743
    /product="caspase-8"
    /note="CDS is reported in Acc#AB038985
    alternative 5'UTR"
    /number=3
    1744..3924
    /number=3
    3925..3970
    /product="caspase-8"
    /note="CDS is reported in Acc#AB038985
    alternative 5'UTR"
    /number=4
    3971..>6502
    /note="intron is approximately 5.9 kbp long"
    /number=4
    1718 a 1361 c 1504 g 1919 t
    ORIGIN
    Query Match      66.5%; Score 500.6; DB 9; Length 6502;
    Best Local Similarity 90.2%; Pred. No. 3.8e-145;
    Matches 665; Conservative 0; Mismatches 44; Indels 28; Gaps 11;

QY      13 TATTGAAGTAAAGAACTTCTTCTGAGACCTTCCACCCCTTCCCTGCTGAGCA 72
DB      1622 TATTGAAGTAAAGAACTTCTTCTGAGACCTTCCACCCCTTCCCTGCTGAGCA 1681

```

REFERENCE
AUTHORS
TITLE
JOURNAL

University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 181150)
Waterston, R.H.

Direct Submission
Submitted (20-APR-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

6 (bases 1 to 181150)
Waterston, R.H.

Direct Submission
Submitted (07-JUN-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
7 (bases 1 to 181150)

REFERENCE
AUTHORS
TITLE
JOURNAL

Waterston, R.H.

Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
8 (bases 1 to 181150)
Waterston, R.

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission

Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jun 7, 2001 this sequence version replaced g1.13677176.

COMMENT

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: M0GSC

Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu

----- Summary Statistics

Center project name: H_NH0575C06

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPO1-11 human BAC library was made from the blood of one male
donor, as described by Oseegawa, K., Woon, P.Y., Zhao, B., Frenken, E.,
Tateno, M., Catanesi, J.V. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)

VECTOR: PBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-156B7; the clone sequenced
to the right is RP11-536I18, 200 bp overlap. Actual start of this
clone is at base position 1 of RP11-575C6; actual end is at base
position 180956 of RP11-575C6.

The sequence from base position 134614 to 134714 is derived from
PCR product of RP11-575C6 BAC DNA.

FEATURES
source The clone RP11-575C6 contains a transposon in the vector.
location/Qualifiers

1..181150	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-575C6"
	/clone_1kb="RPO1-11"
1..354	
	/rpt_family="L1"
91..95	
	/note="similar to Homo sapiens EST BG495251 (NID:g13456766)"
124..153	
	/rpt_family="(TG)n"
193..228	
	/rpt_family="(T)n"
330..1055	
	/rpt_family="L1"
553..620	
	/rpt_family="T-rich"
1056..1363	
	/rpt_family="Alu"
1210..1243	
	/rpt_family="(T)n"
1228..1246	
	/note="match to EST AW444732 (NID:g9586494)"
1364..2264	
	/rpt_family="L1"
3338..3350	
	/rpt_family="MIR"
3739..3866	
	/rpt_family="MIR"
3971..4133	
	/note="similar to Homo sapiens EST AU117593 (NID:g10932573)"
4041..4495	
	/note="similar to Homo sapiens EST AI499220 (NID:g4391202) to08f03.x1"
4053..4639	
	/note="similar to Mus musculus EST BF163999 (NID:g11044285)"
4092..4850	
	/note="similar to Homo sapiens EST AL528010 (NID:g12791503)"
4128..4340	
	/note="match to EST AV760501 (NID:g10918349)"
4184..4742	
	/note="match to EST AI052250 (NID:g3308241) 0221c07.x1"
4339..4903	
	/note="similar to Homo sapiens EST AU117593 (NID:g10932573)"
4485..4491	
	/note="similar to Homo sapiens EST BE788113 (NID:g10209311)"
4657..5348	
	/note="similar to Homo sapiens EST AL563032 (NID:g12912044)"
4712..5400	
	/note="match to EST AU13136 (NID:g109922290)"
4780..5582	
	/note="similar to Homo sapiens EST BG164297 (NID:g12671000)"
4963..5582	
	/note="similar to Homo sapiens EST AU124988 (NID:g10949704)"
5106..5563	
	/note="similar to Mus musculus EST A116325 (NID:g3516649) ue37c04.y1"
5169..5575	
	/note="similar to Homo sapiens EST AU144381 (NID:g11005902)"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:07:09 ; Search time 2674.33 Seconds
(without alignments)
8194.363 Million cell updates/sec

Title: US-09-477-082-2
Perfect score: 753
Sequence: 1 aatagaccgcgattgaaa.....tacacrgtcttttaacct 753

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: **
1: gb_da: **
2: gb_htg: **
3: gb_in: **
4: gb_om: **
5: gb_ov: **
6: gb_pat: **
7: gb_ph: **
8: gb_pl: **
9: gb_pr: **
10: gb_ro: **
11: gb_sts: **
12: gb_sy: **
13: gb_un: **
14: gb_vl: **
15: em_da: **
16: em_fun: **
17: em_hum: **
18: em_in: **
19: em_mu: **
20: em_om: **
21: em_or: **
22: em_ov: **
23: em_pat: **
24: em_ph: **
25: em_pl: **
26: em_ro: **
27: em_sts: **
28: em_un: **
29: em_vl: **
30: em_htg_hum: **
31: em_htg_inv: **
32: em_htg_other: **
33: em_htg_mus: **
34: em_htg_pln: **
35: em_htg_rtd: **
36: em_htg_man: **
37: em_htg_vrt: **
38: em_sy: **
39: em_htgo_hum: **
40: em_htgo_mus: **
41: em_htgo_other: **

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	529.8	70.4	179216	2	AC074016
2	529.8	70.4	181150	9	AC007256
3	500.6	66.5	6502	9	AB038980s2
4	488.4	64.9	982	9	AF210257
5	123.6	16.4	2503	9	AF422925
6	123.6	16.4	2887	6	AR211527
7	123.6	16.4	2887	6	AR134419
8	123.6	16.4	2887	9	HSBACH1
9	121.8	16.2	777	9	AF422929
10	83.8	11.1	7218	6	166494
11	62.2	8.3	128941	2	AC110348
12	45	6.0	146383	2	AC116367
13	45	6.0	146383	2	AL359032-4
14	44.6	5.9	61246	2	AC064809
15	44.6	5.9	145763	2	AL390844
16	44.6	5.9	163677	9	AL390844
17	44	5.8	237727	2	AC099291
18	43.4	5.8	168975	2	AC117170
19	43	5.7	124156	9	AL596132
20	43	5.7	183617	2	AC007524
21	42.8	5.7	128941	2	AC110348
22	42.4	5.6	177407	2	AC091553
23	41.4	5.5	143394	2	AC103197
24	41.4	5.5	171749	2	AC112440
25	41.4	5.5	182360	2	AC107184
26	41.4	5.5	183151	2	AC112825
27	41.2	5.5	125020	9	AF429315
28	41	5.4	44429	2	AC083762
29	41	5.4	56342	2	AC103270
30	41	5.4	230290	2	AC101859
31	41	5.4	303091	2	AC084799
32	41	5.4	176691	2	AC097904
33	40.6	5.3	106582	2	AC114068
34	40.2	5.3	110000	2	AC084053-1
35	40.2	5.3	182539	2	AC110338
36	40.2	5.3	157665	10	AC069018
37	40	5.3	161396	2	AC095069
38	40	5.3	173418	2	AC110354
39	39.8	5.3	969	11	CNS066DL
40	39.8	5.3	128124	2	AC113915
41	39.8	5.3	142333	2	AC107270
42	39.8	5.3	167470	2	AP005545
43	39.8	5.3	171457	2	AC095247
44	39.6	5.3	196912	10	AL513346
45					AL513346 Mouse DNA

ALIGNMENTS

RESULT 1
AC074016 179216 bp DNA linear HTG 09-AUG-2001
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-155P18, WORKING DRAFT
AC074016
ACCESSION AC074016.5 GI:15144340
VERSION
KEYWORDS HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 179216)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone

Publication No. US20030082605A1
GENERAL INFORMATION:
APPLICANT: OLEX, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 1003529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
2000-04-06

FEATURE:
NAME/KEY: unsure
LOCATION: (15565, 15570, 15574, 15580, 15585..15586, 15589..15590, 17255)
US-10-239-676-53

Query Match
Best Local Similarity 4.9%; Score 32.6; DB 9; Length 17421;
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 582 TTACGCTGGAGGAGTGTTCACAGGTTCCCTCTTATCTTTGTTGTTT 641
DB 13184 TTTAGAGAGAGAGAGAGTGTATTAAGTGTGTTTGGAGTTTATTTT 13243
QY 642 CGAGCCATGGGGGTAAATTAAGCGCT 668
DB 13244 TTGGGTGGAGGTTAAATATGAGGT 13270

RESULT 8
US-08-755-235-3/c
Sequence 3, Application US/08755235
Publication No. US20030059423A1
GENERAL INFORMATION:
APPLICANT: Stern, David M.
APPLICANT: Wu, Jun
TITLE OF INVENTION: METHOD FOR TREATING SYMPTOMS OF DIABETES
FILE REFERENCE: 0575/50159
CURRENT APPLICATION NUMBER: US/08/755,235
CURRENT FILING DATE: 1996-11-22
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 1405
TYPE: DNA
ORGANISM: Human
US-08-755-235-3

Query Match
Best Local Similarity 4.8%; Score 32.4; DB 7; Length 1405;
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 176 AGCCACAGGGCTACTTACCCAGTCCGGGAGGAGGAGGAGGCTGTCTGACTT 235
DB 839 AGGGGAGGGGACACACCTTCATCCAGTGTGAGGAGGAGGCTGGGAGGACTT 780
QY 236 CAGTGTGAGGTTT 249
DB 779 CACAGTCAAGGTT 766

RESULT 9
US-08-905-709-3/c
Sequence 3, Application US/08905709
Patent No. US20010039256A1
GENERAL INFORMATION:
APPLICANT: Stern, David
APPLICANT: Schmidt, Ann M.
TITLE OF INVENTION: A METHOD TO PREVENT ACCELERATED
TITLE OF INVENTION: ATHEROSCLEROSIS USING (SRAGE) SOLUBLE RECEPTOR FOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,709
FILING DATE: 05-AUG-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52876
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1405 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-905-709-3

Query Match
Best Local Similarity 4.8%; Score 32.4; DB 8; Length 1405;
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 176 AGCCACAGGGCTACTTACCCAGTCCGGGAGGAGGAGGAGGCTGTCTGACTT 235
DB 839 AGGGGAGGGGACACACCTTCATCCAGTGTGAGGAGGAGGCTGGGAGGACTT 780
QY 236 CAGTGTGAGGTTT 249
DB 779 CACAGTCAAGGTT 766

RESULT 10
US-09-822-830A-512
Sequence 512, Application US/09822830A
Patent No. US20020142952A1
GENERAL INFORMATION:
APPLICANT: Genetics Institute, Inc.
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Agostino, Michael J.
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakari
APPLICANT: Graham, James R.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6402
CURRENT APPLICATION NUMBER: US/09/822,830A
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,604
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 631
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 512
LENGTH: 1426
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-830A-512

Query Match
Best Local Similarity 4.8%; Score 32.2; DB 10; Length 1426;
Matches 88; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 481 TGTTCAGTACATCTCTGTTGCTTTAGAGTAAGTTTACCTGCATCTCTCT 540
DB 903 TGGCACAATGATTTATCTTACCTGCTGATGTGATGATGTCTGCTCTCA 962
QY 541 GTGGTGAAGTTTCTCTCTCTCGAGACAGATTCGCTTTACGCTGAGGGAAGT 600
DB 963 GTTGTATTATTTACATTTGTCACGTTGTTGAAGAAATTTAACAATGATTAAGTCTC 1022

;; TITLE OF INVENTION: 97 Human secreted proteins
;; FILE REFERENCE: P2028p1
;; CURRENT APPLICATION NUMBER: US/09/892,877
;; CURRENT FILING DATE: 2001-06-28
;; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
;; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
;; NUMBER OF SEQ ID NOS: 461
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 103
;; LENGTH: 704
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (287)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-892-877-103

Query Match 4.9%; Score 32.6; DB 9; Length 704;
Best Local Similarity 52.6%; Pred. No. 3.4;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 149 TTTTCTGACGATCTACCCGATTTGACGCACAGGGCTGACTTTACCCAGTCGGCGGG 208
DB 525 TATACTGGAGGCGACTATAGGTAGGTAAATCTTTATTATTATTATTGAGAC 584
QY 209 AGGAGGAGAGGGCTGCTGTGACTGAGTGTGAGTTGATCAAGGCAAGGGAAC 268
DB 585 AGGAGGGCTCTTGTCTTTGTTTCAGACTGGAGTGGTGTGATCATGTGCTCATTTGCAAC 644
QY 269 TTCTATTCACGAC 283
DB 645 TTGAACCTCTGGGC 659

RESULT 6
US-09-948-783-103
;; Sequence 103, Application US/09948783
;; Publication No. US20030100051A1
;; GENERAL INFORMATION:
;; APPLICANT: Ruben et. al.
;; TITLE OF INVENTION: 97 Human secreted proteins
;; FILE REFERENCE: P2028p2
;; CURRENT APPLICATION NUMBER: US/09/948,783
;; CURRENT FILING DATE: 2001-09-10
;; PRIOR APPLICATION NUMBER: 60/231,846
;; PRIOR FILING DATE: 2000-09-11
;; PRIOR APPLICATION NUMBER: 09/892,877
;; PRIOR FILING DATE: 2001-06-28
;; PRIOR APPLICATION NUMBER: 09/437,658
;; PRIOR FILING DATE: 1999-11-10
;; PRIOR APPLICATION NUMBER: PCT/US99/09847
;; PRIOR FILING DATE: 1999-05-06
;; PRIOR APPLICATION NUMBER: 60/085,093
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,094
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,105
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,180
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085,927
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,906
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,924
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,922
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,921
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,923
;; PRIOR FILING DATE: 1998-05-18

;; PRIOR APPLICATION NUMBER: 60/085,925
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,928
;; PRIOR FILING DATE: 1998-05-18
;; PRIOR APPLICATION NUMBER: 60/085,920
;; PRIOR FILING DATE: 1998-05-18
;; NUMBER OF SEQ ID NOS: 465
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 103
;; LENGTH: 704
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: SITE
;; LOCATION: (287)
;; OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-103

Query Match 4.9%; Score 32.6; DB 9; Length 704;
Best Local Similarity 52.6%; Pred. No. 3.4;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 149 TTTTCTGACGATCTACCCGATTTGACGCACAGGGCTGACTTTACCCAGTCGGCGGG 208
DB 525 TATACTGGAGGCGACTATAGGTAGGTAAATCTTTATTATTATTATTGAGAC 584
QY 209 AGGAGGAGAGGGCTGCTGTGACTGAGTGTGAGTTGATCAAGGCAAGGGAAC 268
DB 585 AGGAGGGCTCTTGTCTTTGTTTCAGACTGGAGTGGTGTGATCATGTGCTCATTTGCAAC 644
QY 269 TTCTATTCACGAC 283
DB 645 TTGAACCTCTGGGC 659

RESULT 7
US-10-239-676-53
;; Sequence 53, Application US/10239676
;; Publication No. US20030082609A1
;; GENERAL INFORMATION:
;; APPLICANT: OLEK, Alexander
;; APPLICANT: PIEPENROCK, Christian
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;; FILE REFERENCE: 5013.1003
;; CURRENT APPLICATION NUMBER: US/10/239,676
;; CURRENT FILING DATE: 2002-09-24
;; PRIOR APPLICATION NUMBER: PCT/EP01/03968
;; DE 10019058.8
;; DE 10019173.8
;; DE 10032529.7
;; DE 10043826.1
;; PRIOR FILING DATE: 2001-04-06
;; 2000-04-06
;; 2000-04-07
;; 2000-06-30
;; 2000-09-01
;; NUMBER OF SEQ ID NOS: 228
;; SEQ ID NO 53
;; LENGTH: 17421
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (515, 1325, 1345, 3705, 3707, 5320, 5410, 5417, 5420, 5426)
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (7028, 7082, 7179, 8249, 8252, 8254, 8269, 8275, 8278, 8279)
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (9222, 9233, 9261, 12655, 12689, 12792, 13398, 14469, 15562, 15563)

```

RESULT 3
US-09-918-995-12139/c
Sequence 12139, Application US/09918995
Publication No. US20090073623A1
GENERAL INFORMATION:
APPLICANT: Hysco, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 200411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30

```

Query Match	Similarity	4.9%	Score 33.8	DB 9	Length 7010
Best Local	Similarity	53.0%	Pred. No. 10		
Matches	Conservative	0	Mismatches	62	Indels 0; Gaps 0;
QY	527	CTGCAGTTCCTTCGTGTGTAAGTTTCCTTCCTCTCGAGACACAGATTCGCTTAC	586		
Db	2962	CTGACATTCCTCCCAAGATTGCGACGATGATCTTTGATTTGCTTCCAGATGACACTTAG	2903		
QY	587	GCTGAGGGAAGTGTTCACACAGGTCCTCCCTTTATCTTTGTGTGTTTTTTTCGAGC	646		
Db	2902	TTTGGAGCAATTGGAGTGTGTGTTCTCTTTCTTTTTTTTTTTTTTTTTTTTATACACA	2843		
QY	647	CATGGGGGTAA	658		
Db	2842	AGACGTGGTAA	2831		

RESULT 5
US-09-892-877-103
; Sequence 103, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruden et. al.

[illegible]

Db	442	TCAGATTGGAGAGTTCAGAAAATGAAAATGAGGTAAGTTGTCGTGAGAG	383
QY	106	CAGGCTAAGCCAGTACGATGAAACGACACTTCCTCTTTTCTGACGATCTA	165
Db	382	ATGGGTATATACATTTATAAAGAAATATTTTCTTTAATTTCTCTACTATTTA	323
QY	166	CCGCAATTCAGCCACAGGGGCTGACTT	192
Db	322	TCACCATATTTTCCCACTTTTACAT	296

Search completed: July 10, 2003, 22:32:56
 Job time : 55 secs


```

      FILING DATE:
      CLASSIFICATION:
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/840,146
      FILING DATE: 11-APR-1997
      ATTORNEY/AGENT INFORMATION:
      NAME: Arnold, Beth E.
      REGISTRATION NUMBER: 35,430
      REFERENCE/DOCKET NUMBER: MTA-018.01
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 617-832-1000
      TELEFAX: 617-832-7000
      INFORMATION FOR SEQ ID NO: 16:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 554 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      US-09-360-220-16

Query Match      4 5%: Score 30.4; DB 3; Length 554;
Best Local Similarity 55.6%; Prcd.No.1.6;
Matches 79; Conservative 0; Mismatches 61; Indels 2; Gaps 1

      QY      517 AAGATTACCCGACAGTTCCTTCGTGTGAGAGTTTCTCTTCGAGAGACAGATT 576
      DB      290 AATTGGGTAGAGAGACACTGCTGCTAGAGAGCTGCTCTACCTGCAGAGAACCTG--G 233
      QY      577 CCGCCTTAGCGTGGAGGAGAGTGTTCACAGGTTCTCCTCTTATCTTTTGCTT 636
      DB      232 AAGGACAGACGCTGCGAGGCCAGAGTAGCCCTGCTCCTCGAGATTTCCTTTAACCTTC 173
      QY      637 TTTTGGACGATGGGGTTAA 658
      DB      172 CTTATGGCTCATAGGGTTCA 151

      RESULT 13
      US-08-840-146-1/c
      Sequence No. 1, Application US/08840146
      Patent No. 6037173
      GENERAL INFORMATION:
      APPLICANT: Glucksmann, M. Alexandra
      TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS AND
      TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR DISEASES INVOLVING TRBP
      NUMBER OF SEQUENCES: 20
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: FOLEY, HONG & ELLIOT LLP
      STREET: One Post Office Square
      CITY: Boston
      STATE: MA
      COUNTRY: USA
      ZIP: 02109-2170
      COMPUTER READABLE FORM:
      MEDIUM TYPE: floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patent Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/840,146
      FILING DATE: 11-APR-1997
      CLASSIFICATION: 435
      ATTORNEY/AGENT INFORMATION:
      NAME: Arnold, Beth E.
      REGISTRATION NUMBER: 35,430
      REFERENCE/DOCKET NUMBER: MTA-018.01
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 617-832-1000
      TELEFAX: 617-832-7000
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 2241 base pairs

```

```

CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/633,148
FILING DATE: 16-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY ESQ., MATTHEW B.
REGISTRATION NUMBER: 39,787
REFERENCE/DOCKET NUMBER: 014618-005600US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 957 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-633-148-3

Query Match
Best Local Similarity 66.2%; Score 31.2; DB 2; Length 957;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 182 AGGCGTACTTTACCCAGTCGGCGGAGAGAGAGAGGCGTGTGACTTCAGTGC 241
Db 766 AGGCGACACCATCTTCATCCAGTGTGAGAGAGAGGCGTGGCGAGGACTTCACAGG 707
QY 242 TGAGGTTT 249
Db 706 TCAGGGTT 699

RESULT 9
US-08-633-148-1/c
Sequence 1, Application US/08633148
Patent No. 5864018
GENERAL INFORMATION:
APPLICANT: MORSE, MICHAEL J.
APPLICANT: NAGASHIMA, MARIKO
APPLICANT: HOLLANDER, DORIS A.
TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/633,148
FILING DATE: 16-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY ESQ., MATTHEW B.
REGISTRATION NUMBER: 39,787
REFERENCE/DOCKET NUMBER: 014618-005600US

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-633-148-1

Query Match
Best Local Similarity 66.2%; Score 31.2; DB 2; Length 1023;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 182 AGGCGTACTTTACCCAGTCGGCGGAGAGAGAGGCGTGTGACTTCAGTGC 241
Db 832 AGGCGACACCATCTTCATCCAGTGTGAGAGAGGCGTGGCGAGGACTTCACAGG 773
QY 242 TGAGGTTT 249
Db 772 TCAGGGTT 765

RESULT 10
US-09-457-066-42/c
Sequence 42, Application US/09457066
Patent No. 6432673
GENERAL INFORMATION:
APPLICANT: Gao, Zeren
APPLICANT: Hart, Charles E.
APPLICANT: Piddington, Christopher S.
APPLICANT: Sheppard, Paul O.
APPLICANT: Shoemaker, Kimberly E.
APPLICANT: Gilbertson, Debra G.
TITLE OF INVENTION: GROWTH FACTOR HOMOLOG ZVEGF3
FILE REFERENCE: 98-60
CURRENT APPLICATION NUMBER: US/09/457,066
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 50
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 42
LENGTH: 3571
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1049)...(2086)
US-09-457-066-42

Query Match
Best Local Similarity 58.9%; Score 30.8; DB 4; Length 3573;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 548 AGTTTCCTCTTCTCGGAGACAGATTCGCTTACGCTGAGAGAGTGTTCAC 607
Db 2372 ATTGTCTCTGTGTCGAGAGACCATTCGTGCGCTTCCTAGAGGAGGCTTGTG 2313
QY 608 AGGTCCTCCCTCTTATCTTTGTTT 637
Db 2312 ATTGACTCTCTTGTGTCCTTGCTTTT 2283

RESULT 11
US-08-840-146-16/c
Sequence 16, Application US/08840146
Patent No. 6037173
GENERAL INFORMATION:
APPLICANT: Glucksmann, M. Alexandra
TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND METHODS AND
TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR DISEASES INVOLVING TRBP

```

[illegible]

[illegible]

```

: RESULT 4
: PCT-US93-05000-32
: Sequence 32, Application PC/TUS9305000
: GENERAL INFORMATION:
: APPLICANT: MITOXIX
: TITLE OF INVENTION: D-Type Cyclin and Uses Related thereto
: NUMBER OF SEQUENCES: 42
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: US
: ZIP: 02173
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/05000
: FILING DATE: 19930525
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/888,178
: FILING DATE: 26-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: CSHL91-02A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 616-861-9540
: INFORMATION FOR SEQ ID NO: 32:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1901 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: PCT-US93-05000-32

```

[illegible]

```

1      RESULT 5
2      US-08-464-517-32
3      Sequence 32 Application US/08464517
4      Patent No. 5869640
5      GENERAL INFORMATION:
6      APPLICANT: BEACH, David H.
7      TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
8      NUMBER OF SEQUENCES: 50
9      CORRESPONDENCE ADDRESS:
10     ADDRESSEE: LAHIVE & COCKFIELD
11     STREET: 60 State Street
12     CITY: Boston
13     STATE: MA
14     COUNTRY: USA
15     ZIP: 02109
16     COMPUTER READABLE FORM:
17     MEDIUM TYPE: Floppy disk
18     COMPUTER: IBM PC compatible
19     OPERATING SYSTEM: PC-DOS/MS-DOS
20     SOFTWARE: ASCII(text)
21     CURRENT APPLICATION DATA:
22     APPLICATION NUMBER: US/08/464,517
23     FILING DATE:
24     CLASSIFICATION: 435
25     PRIOR APPLICATION DATA:
26     APPLICATION NUMBER: US 07/963,308
27     FILING DATE: 16-OCT-1992
28     APPLICATION NUMBER: US 07/888,178
29     FILING DATE: 26-MAY-1992
30     PRIOR APPLICATION DATA:
31     APPLICATION NUMBER: US 07/701,514
32     FILING DATE: 16-MAY-1991
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Matthew P. Vincent
35     REGISTRATION NUMBER: 36,709
36     REFERENCE/DOCKET NUMBER: MTI-004C
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE: (617) 227-7400
39     TELEFAX: (617) 227-5941
40     INFORMATION FOR SEQ ID NO: 32:
41     SEQUENCE CHARACTERISTICS:
42     LENGTH: 2022 base pairs
43     TYPE: nucleic acid
44     STRANDEDNESS: double
45     TOPOLOGY: linear
46     MOLECULE TYPE: DNA (genomic)
47     FEATURE:
48     NAME/KEY: CDS
49     LOCATION: join(1137..1211, 1211..1678, 1680..1790)
50     US-08-464-517-32

```

```

Query Match          5.0%; Score 33.2; DB 2; Length 2022;
Best Local Similarity 51.3%; Pred. No. 0.37;
Matches 77; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY      413  TCTGCCAATCCCTCTCTGTAATGTTGGCATTTGGGCATCTGTCTTCTTAACAGGAA 472
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       539  TCTCTGCTCTAGATCTCTGAATAGCTTAATATACAGGCACTCGCACTGAGCC 598
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      473  ACATTTCTGTGTGAGAGTGCATCTGTCTGCTTTAGAGTAATACCTTGAC 532
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       599  ATATTTTGTGTTGTGTTTGTGTTGTGTTGTGTTGTGTTGTGTTGTGTTGTG 658
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY      533  TTCTCTCTGCTGAGATTCTCTTTCTC 562
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       659  CCCAGCGCTGAGTGCAGTGGCGGCGATCTC 688

RESULT 6
US-08-246-361A-32
; Sequence 32, Application US/08246361A
; Patent No. 5998582
; GENERAL INFORMATION:

```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:21:05 ; Search time 51 Seconds

(Without alignments)
4028,891 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670
Sequence: 1 aagcgcctccaaagacagcatt.....ggggttaataaagcgttt 670

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCUTUS.COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfilest.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.6	10.7	2887	US-08-983-502-14	Sequence 14, Appl
2	71.6	10.7	2887	PCT-US96-10521-14	Sequence 14, Appl
3	49.6	7.4	7218	US-08-233-463-14	Sequence 14, Appl
4	33.2	5.0	1901	PCT-US93-05000-32	Sequence 32, Appl
5	33.2	5.0	2022	US-08-464-517-32	Sequence 32, Appl
6	33.2	5.0	2022	US-08-246-361A-32	Sequence 32, Appl
7	33.2	5.0	2022	US-08-463-772-32	Sequence 32, Appl
8	31.2	4.7	957	US-08-633-148-3	Sequence 3, Appl
9	31.2	4.7	1023	US-08-633-148-1	Sequence 3, Appl
10	30.8	4.6	3573	US-09-457-066-42	Sequence 1, Appl
11	30.4	4.5	554	US-08-840-146-16	Sequence 16, Appl
12	30.4	4.5	554	US-09-360-220-16	Sequence 16, Appl
13	30.4	4.5	2241	US-08-840-146-1	Sequence 1, Appl
14	30.4	4.5	2241	US-09-360-220-1	Sequence 1, Appl
15	30.2	4.5	1457	US-08-460-512-1	Sequence 1, Appl
16	29.8	4.4	18627	US-08-961-527-113	Sequence 113, Appl
17	29.8	4.4	40352	US-08-846-111D-15	Sequence 15, Appl
18	29.8	4.4	40352	US-09-444-077-15	Sequence 15, Appl
19	29.6	4.4	1186	US-09-222-017B-478	Sequence 478, Appl
20	29.6	4.4	2248	US-08-633-237-1	Sequence 1, Appl
21	29.6	4.4	2264	US-08-975-405-1	Sequence 1, Appl
22	29.6	4.4	7218	US-09-167-109-6	Sequence 6, Appl
23	29.6	4.4	7218	US-08-233-463-14	Sequence 14, Appl
24	29.4	4.4	446	US-09-222-575-93	Sequence 93, Appl
25	29.2	4.4	976	US-08-248-839C-5	Sequence 5, Appl
26	29.2	4.4	983	US-08-248-839C-1	Sequence 1, Appl
27	29.2	4.4	1135	US-08-002-024B-7	Sequence 7, Appl

C 28	29.2	4.4	1200	US-08-248-839C-7	Sequence 7, Appl
C 29	29.2	4.4	7680	PCT-US95-09819-6	Sequence 6, Appl
C 30	29.2	4.4	7705	US-08-259-569-16	Sequence 16, Appl
C 31	29.2	4.4	7705	US-08-826-885-16	Sequence 16, Appl
C 32	29.2	4.4	7705	54551582	Patent No. 5455158
C 33	29.2	4.4	7803	US-08-551-356-1	Sequence 1, Appl
C 34	29.2	4.4	7803	PCT-US93-12687-1	Sequence 1, Appl
C 35	29	4.3	8791	US-09-538-414-10	Sequence 10, Appl
C 36	29	4.3	13737	US-09-538-414-10	Sequence 10, Appl
C 37	29	4.3	14113	US-08-992-801-1	Sequence 1, Appl
C 38	29	4.3	14113	US-08-992-801-1	Sequence 1, Appl
C 39	29	4.3	14113	US-09-223-535-1	Sequence 1, Appl
C 40	28.8	4.3	387	US-08-686-878A-52	Sequence 52, Appl
C 41	28.8	4.3	387	US-08-721-924-2	Sequence 2, Appl
C 42	28.8	4.3	387	US-09-175-928-6	Sequence 6, Appl
C 43	28.8	4.3	53526	US-08-658-136-1	Sequence 2, Appl
C 44	28.8	4.3	53577	US-08-658-136-1	Sequence 1, Appl
C 45	28.8	4.3	4411529	US-09-103-840A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-983-502-14
Sequence 14, Application US/08983502
Patent No. 6399327
GENERAL INFORMATION:
APPLICANT: David WALLACH
APPLICANT: Mark P. BOLDIN
APPLICANT: Tanya M. GONCHAROV
APPLICANT: Yuri V. GOLTSEV
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
TITLE OF INVENTION: AND OTHER PROTEINS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Broadway and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/983,502
FILING DATE: 16-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10521
FILING DATE: 14-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 114,615
FILING DATE: 16-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 114,986
FILING DATE: 17-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 115,319
FILING DATE: 14-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 116,568
FILING DATE: 27-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 117,932
FILING DATE: 16-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Broadway, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH-19
TELECOMMUNICATION INFORMATION:

Query Match 5.8%; Score 38.8; DB 17; Length 654;
 Best Local Similarity 50.0%; Pred. No. 24;
 Matches 97; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 332 ATTACTAAATGAGATCGATATAAGCTTTCCATAAAGCATGCGAGCGGCTTT 391
 DB 322 AATATACATTTGATATATAAGAGAGATCTGAAAAGTGAATCCATCATGCGAGCTTT 263
 QY 392 AGTTTCAGCTGCATGATGTTCTGCGACATCCCTCTTGATAGTGGATGGCAT 451
 DB 262 ATTATGACCTTTTACTAATCTGCTCTTAAGTGGCAGATTTACTAAGTGGAGCTTT 203
 QY 452 CTCCTGCTTTAAACAGGAACATTTCTGTCGAGTGCATCTCTGCTCTTTA 511
 DB 202 TGGTTTCTTTTAAACAATCTGCATTTCTTATCTCTCTAGTTTTTTCTTTT 143
 QY 512 GGAGTAAAGTTTAC 525
 DB 142 GCTGATTTATTAAC 129

RESULT 13
 LOCUS AG043660/c 1080 bp DNA linear GSS 01-NOV-2001
 DEFINITION Pan troglodytes DNA, clone: PTB-022A16.R, genomic survey sequence.
 ACCESSION AG043660
 VERSION AG043660.1 GI:16572385
 KEYWORDS GSS.
 SOURCE Pan troglodytes male lymphoblast DNA, clone lib: PTB Chimpanzee Male
 ORGANISM Pan troglodytes
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.
 TITLE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Matanabe, H. and Sakaki, Y.
 JOURNAL BAC end sequences of library PTB
 REFERENCE 2 (bases 1 to 1080)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Matanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan 1-7-22 Suhei-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/), Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the chimpanzee BAC library PTB this BAC end was generated during the Rad process and may have higher chance of clone tracking errors.
 PRIMERS
 Sequencing: M13Rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI
 Location/Qualifiers
 1.1080
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone_lib="PTB-022A16.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"
 BASE COUNT 573 a 55 c 287 g 67 t 98 others
 ORIGIN

Query Match 5.8%; Score 38.6; DB 17; Length 1080;
 Best Local Similarity 49.4%; Pred. No. 26;
 Matches 83; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
 QY 474 CATTTCTTTGAGAGTCACTCTGTTCTGCTTTAGAGTAAGTTACCTGCACT 533

DB 809 CTTTCCCT 750
 QY 534 TCT 553
 DB 749 TTTCT 690
 QY 594 GGAAGTGTTCACAGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 641
 DB 689 TGTCT 642

RESULT 14
 CENSOR/LT/c 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence 17 end of BAC:
 DEFINITION BACR30P10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL068307
 VERSION AL068307.1 GI:4958538
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 TITLE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 JOURNAL Direct Submission
 COMMENT Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqret@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila> and melanogaster BAC library was prepared by Kazuo Osega and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1.1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="BACR30P10"
 /clone_lib="RPCI-98"
 /note="end : 17"
 BASE COUNT 507 a 148 c 112 g 171 t 163 others
 ORIGIN

Query Match 5.8%; Score 38.6; DB 17; Length 1101;
 Best Local Similarity 35.0%; Pred. No. 26;
 Matches 97; Conservative 42; Mismatches 188; Indels 0; Gaps 0;

QY 365 AATAAGCATGTCAGGCGCTGAGTTAGTGCACGTCATGATGTCTGCCACATCC 424
 DB 607 AATAAGCATGTTGAACCCAGCTTCMCAGGCTAAAGATTTTCTTGATGATCWTGT 548
 QY 425 CTCCTCTGATGATGTTGAATGGGACATCTCTCTTAAACAGGAACATTTCTGTT 484
 DB 547 CTTTWTAAAGGGGAGGCTCTTTTCTTGATGATGATGATGATGATGATGATGAT 488
 QY 485 CGAGGATCATCT 544
 DB 487 TTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT 428

```

/clone_1lb="Soares fetal liver spleen INFLS"
/dev="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH103 (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pRTT3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGAGAGATTAATTAAGATCTTTTATTTTATTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pRTT3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo."
BASE COUNT      129 a      76 c      100 g      90 t      3 others
ORIGIN

Query Match      5.8%; Score 39; DB 14; Length 398;
Best Local Similarity 52.2%; Pred. No. 21;
Matches 84; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 469 GGAACATTCTTCTGAGTGAATCATCTCTGCTTGTAGAGTAAGTTACCC 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 341 GAAATCTTTTTCCTTNCCTTGAAGAAAGCTGTAACTATTAGATTTTATA 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 529 GCAGTCTCTCTGCTGGAAGTTTCTCTCTCTCGAGACAGATTCGCTTACGC 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 281 GGACCTTTTAAATTTTAAATTTTGTGAGCAGAGGCTCCACCTGTCGCCAGGC 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 589 TGGAGGAGAGTGTTCACAGGTTCTCTCTCTTATCTT 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 221 TGGAGTCACTGCTTCAACACACAGCTTACCTTACCTT 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
CNS001W9      423 bp      DNA      linear      GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence 17 end of BAC:
DEFINITION      BACR37D10 of RPCR-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION      AL075520
VERSION      AL075520.1 GI:4955098
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster.
ORGANISM      Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 423)
AUTHORS      Genoscope.
TITLES      Direct Submission
JOURNAL      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazuhiro Oseegawa and
            Aaron Mammeter in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCR-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp; the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
            Location/Qualifiers
            1..423
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="BACR37D10"

```

```

/clone_1lb="RPCI-98"
/dev="end : 17"
/dev_stage="29 c"
/lab_host="6 g"
/note="32 t 53 others
BASE COUNT      303 a      29 c      6 g      32 t      53 others
ORIGIN

Query Match      5.8%; Score 38.8; DB 17; Length 423;
Best Local Similarity 42.5%; Pred. No. 24;
Matches 79; Conservative 20; Mismatches 87; Indels 0; Gaps 0;

QY 476 TTCTCTGCGAGTATCTCTGCTGCTTTAGAGTAAGTTACCTGAGCTC 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 275 TTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 CTCTGCTGAGAGTTTCTCTCTCTCTCTGAGACAGATTCGCTTACGCTGAGG 595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 215 TTGATTTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 596 AAGTCTTTCACAGGTTCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTT 655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 TTTTGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 656 TAAAT 661
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 TATATA 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
AO528592      654 bp      DNA      linear      GSS 18-MAY-1999
LOCUS      RPCR-11-360P2.TV RPCI-11 Homo sapiens genomic clone RPCR-11-360P2,
DEFINITION      DNA sequence.
ACCESSION      AO528592
VERSION      AO528592.1 GI:4840705
KEYWORDS      GSS.
SOURCE      Homo sapiens
            human.
ORGANISM      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 654)
AUTHORS      Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
TITLES      Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building (1997)
JOURNAL      Unpublished (1997)
COMMENT      Other-GSSs: RPCR-11-360P2.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbeetlgr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieterdejong@med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet. cs (info@resgen.com). BAC end search page:
            http://www.lifg.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.
            Seg primer: 17
            Class: BAC ends.
            Location/Qualifiers
            1..654
            /organism="Homo sapiens"
            /db_xref="GBS:7638217"
            /db_xref="taxon:9606"
            /clone="RPCI-11-360P2"
            /clone_1lb="RPCI-11"
            /sex="Male"
            /cell_type="Lymphocytes"
            /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
            RPCI11 Human Male BAC Library"
BASE COUNT      213 a      116 c      114 g      211 t
ORIGIN

```


RESULT 6
LOCUS BQ941968
DEFINITION BQ941968 940 bp mRNA linear EST 21-AUG-2002
ACCESSION AGENCOURT_8835770 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6421760
VERSION BQ941968
KEYWORDS BQ941968.1 GI:22357446
SOURCE EST.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL NIH-MGC <http://mgc.nci.nih.gov/>
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9abs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LICM2597 row: k column: 09
High quality sequence start: 13
High quality sequence stop: 625.
Location/Qualifiers
1. 940
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6421760"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

BASE COUNT 281 a 236 g 223 t
ORIGIN
Query Match 6.3%; Score 42.4; DB 14; Length 940;
Best Local Similarity 89.1%; Pred. No. 2.9; Mismatches 6; Indels 1; Gaps 1;
Matches 57; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
OY 555 TCTTCTCTGAGAGACGATTCGCTTACGCTGAGGAGAGTCTTTTACAGGTCCT 614
DB 1 TCTTCTCTGAGAGACGATTCGCTTACGCTGAGGAGAGTCTTTTACAGATTTAT 59
OY 615 CCTC 618
DB 60 ATTC 63
RESULT 7
LOCUS CNS04NCO
DEFINITION CNS04NCO 921 bp DNA linear GSS 21-MAY-2000
Tetradon nigriviridis genome survey sequence pUC-ori end of clone
122009 of library G from Tetradon nigriviridis, genomic survey
sequence.
ACCESSION AL298521
VERSION AL298521.1 GI:8037101
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigriviridis.
ORGANISM Tetradon nigriviridis

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetradon.
1 (bases 1 to 921)
Roest-Crolius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Filames,C., Winkler,P., Brotlier,P., Queller,F.,
Saurin,M. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigriviridis DNA sequence
Unpublished
2 (bases 1 to 921)
Roest-Crolius,H., Jallou,O., Dasilva,C., Filames,C., Fisher,C.,
Bouneau,L., Billault,A., Queller,F., Saurin,M., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigriviridis
Unpublished
3 (bases 1 to 921)
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigriviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetradon>.
Location/Qualifiers
1. 921
/organism="Tetradon nigriviridis"
/db_xref="taxon:99883"
/clone="122009"
/clone_lib="G"
/note="Genoscope sequence ID : COB6122AH05SP1-end :
pUC-ori"

BASE COUNT 253 a 179 c 182 g 184 t 123 others
ORIGIN
Query Match 6.2%; Score 41.6; DB 17; Length 921;
Best Local Similarity 36.2%; Pred. No. 4.6; Mismatches 87; Indels 0; Gaps 0;
Matches 71; Conservative 38; Mismatches 87; Indels 0; Gaps 0;
OY 475 ATTCTGTTCGAGAGAGATCTGCTTCCTTAGAGAGAAAGTTACCTGCACAT 534
DB 483 AGTCTGCTGATGATGAGTACGTTTCTTATGTTTMMKKAAKCKYKTC 424
OY 535 CCTCTGTGAGTGAAGTTTCTCTCTCGAGAGACGAGATTCCTTACGCTGAGG 594
DB 423 CKYTTCTGTTAKKCTTTTCTTTTGGKKTAAAGKKTWTCCTTTTATTTTMMG 364
OY 595 GAAGTGTTCACAGGTTCTCTCTTATCTTTGTTTGTGAGGACATGAGGG 654
DB 363 CCMTKTKTCTTWTGTTTGMKRYTWMCCATTTTCTGTATGKATGCTGCATGKCTC 304
OY 655 TTAATAAAGCCCTT 670
DB 303 YTTATACAAAGCKWT 288
RESULT 8
LOCUS CNS009VU
DEFINITION CNS009VU 792 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence Tenz end of BAC #
BACR20D15 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053997
VERSION AL053997.1 GI:4935574
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 792)

Db	62	GAAGTTTCCTCCTCTCGGAGACACAAATTCCTGCCTTTCGTGGAGGAATCTTTTC	121
OY	606	ACAGTGTCCTCC	618
Db	122	ACAGATTATATTC	134
RESULT 4			
LOCUS	B1824550	804 bp	mRNA
DEFINITION	603035156F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176548 5',		linear EST 04-OCT-2001
ACCESSION	B1824550		
VERSION	B1824550		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 804)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LRAM11439 row: 0 column: 13 High quality sequence stop: 620. location/Qualifiers 1..804 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5176548" /clone_11b="NIH_MGC_115" /lab_host="DH10B" /note="Organ: pooled brain, lung, testis; Vector: PCWV-SPORE; site_1: NotI; site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 Kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."		
FEATURES			
SOURCE			
BASE COUNT	226 a	187 c	209 g
ORIGIN		182 t	
Query Match	14.1%;	Score 94.8;	DB 13; Length 804;
Best Local Similarity	86.6%;	Pred. NO. 1.9e-13;	
Matches	116;	Conservative	0; Mismatches 17; Indels 1; Gaps 1;
OY	518	AGATTACCCTGCAGTCTCTCTGTGTGTAAGATTTCTTTCTCTCGAG-ACCAGAT	576
Db	3	AAGTTTACCTGCAGTCTCTCATGTGGAGAATTTCTCTCTCGAACAACAGAT	62
OY	577	CTGCCCTTAGCCGTGGAGGAAGTCTTTTCAAGAGTCTCCCTTTATCTTTGTGTT	636
Db	63	CTGCCCTTTCGTCTGAGAGAGATATATCACAGGCTCTCCCTTTATCTTTAGTGTT	122
OY	637	TTTTTCAGCCATG	650
Db	123	TTGTTCAAGCCCTG	136
RESULT 5			

LOCUS	CSNS17GS	1101 bp	DNA	linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence S6 end of BAC BACN17J04 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL108022				
VERSION	AL108022.1 GI:5628326				
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryoeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)				
REFERENCE	Genoscope. Direct Submission Submitted (22-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 J1006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end and sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (DrosBAC) was made by Alain Billard at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.				
FEATURES	Location/Qualifiers				
SOURCE	1..1101 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone="BACN17J04" /clone_1fb="DrosBAC" /plasmid="pBelosBAC11" /note="end : SP6"				
BASE COUNT	268 a 126 c 100 g 195 t 412 others				
ORIGIN					
Query Match	7.2%; Score 48.4; DB 17; Length 1101;				
Best Local Similarity	21.9%; Pred. No. 0.089;				
Matches	98; Conservative 144; Mismatches 209; Indels 0; Gaps 0;				
QY	222 CTGCTGTGACTCTAGCTGCTGAGCTTTGATCAAGCGCAAGGAACTTCCTATTCACG 281				
DB	594 YTTTCTTCTGTTTWWAWTTTWWAMCCCTTATGCGGGRAMWTTTTHSCGACSCMNATVHC 653				
QY	282 ACCCTTTGCAAGAAGATGCGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 341				
DB	654 AAHTWRTAYMDCBANVTTTHTAVABAAGSTTAATAAAMVAAAKTKTGTAAACAMAWA 713				
QY	342 GGAATCAGTAAATGCTTCCATTAAGCATGTCACGCTCGGGCTTTGTTTGACG 401				
DB	714 TYWMMVAATTKTYTAAHMHMAAMCTGYSCYKSTGTGTCBSYKVBGTITGTS 773				
QY	402 TCCATGATTTCTGCGCACATCCCTCTTCTGAATGGTGGATGGGATCTGTCTGCT 461				
DB	774 SCTTYVAHAMDSISKYTCGYCTTTTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 833				
QY	462 TTAACAGGAATTCCTGTTGAGAGTACATCTCTGCTGCTTTAGAGTAAGT 521				
DB	834 TYCBBBGGGSKTKTCTTTTTCYCCYCBSSBYHHYCYCYCBYCYTYTCDPRAWMTS 893				
QY	522 TTACCCGCAATGCTCTGCGGAGAGTTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 581				
DB	894 TKBKTKTKTKTKTKTKTKSKKKKKKKKTSBKTYBVBGVBGGCTSBSSCTVTCYBC 953				
QY	582 TTACAGCTGAGGAGAGTGTTTACAGAGTCTCTCTCTTTATCTTTTGTGTTT 641				
DB	954 YCCYTTCSTTKTTTKYATBTBTTTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 1013				
QY	642 CGAGCCATGGGGTTAATAAGCGCT 668				
DB	1014 BBYKSSGYSCKSSBCKBYBGCT 1040				

FEATURES
source

Location/Qualifiers
1. 522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686E18155"
/clone_1lb="586 (synonym: hlc3)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SflIA; Site_2: SflIB;
CDNA-collection"

BASE COUNT 120 a 116 c 133 g 149 t 4 others

ORIGIN

Query Match 22.6%; Score 151.6; DB 9; Length 522;
Best Local Similarity 96.2%; Pred. No. 16-27; Indels 0; Gaps 0;
Matches 154; Conservative 0; Mismatches 6;

QY 491 AGTCATCTCTGTCCTGCTTGGAGTAAGTTTACCTGCGAGTCTCTGCTGGAAGT 550
|||||
DB 1 AGTCATCTCTGTCCTGCTTGGAGTAAGTTTACCTGCGAGTCTCTGCTGGAAGT 60
|||||

QY 551 TTCTCTCTCTGCTGCGAGACCAAGTCTGCTTACGCTGAGGAGGAGTCTTTCACAGG 610
|||||

DB 61 TTCTCTCTCTGCTGCGAGACCAAGTCTGCTTACGCTGAGGAGGAGTCTTTCACAGG 120
|||||

QY 611 TTCTCTCTCTCTTATCTTTGTTGTTTTCGAGCCATG 650
|||||

DB 121 TTCTCTCTCTTATCTTTGTTGTTTTCGAGCCCTG 160
|||||

RESULT 2
AL702116 421 bp mRNA linear EST 22-MAR-2002
LOCUS DKFZp686E18155_r1 686 (synonym: hlc3) Homo sapiens CDNA clone
DEFINITION DKFZp686E18155 5', mRNA sequence.
ACCESSION AL702116
VERSION AL702116.1 GI:19685471
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 421)
Bloecker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann
S. (Bloecker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and
Wiemann, S.)
Unpublished (1999)
JOURNAL
COMMENT Contact: Bloecker H
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ): Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBR (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the CDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686E18155) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686E18155"
/clone_1lb="586 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SflIA; Site_2: SflIB;
CDNA-collection"

BASE COUNT 95 a 89 c 113 g 124 t

ORIGIN

Query Match 19.7%; Score 132; DB 9; Length 421;
Best Local Similarity 96.4%; Pred. No. 8-9e-23; Indels 0; Gaps 0;
Matches 135; Conservative 0; Mismatches 5;

QY 511 AGGAGTAAGTTTACCTGCGAGTCTCTGCTGAGAGTTTCTCTTCTCGGAGAC 570
|||||

DB 2 AGGAGTAAGTTTACCTGCGAGTCTCTGCTGAGAGTTTCTCTTCTCGGATAC 61
|||||

QY 571 CAGATTCGCTTACGCTGAGGAGAGTTTTCACAGGTTCTCTCTTATCTTT 630
|||||

DB 62 CAGATTCGCTTACGCTGAGGAGAGTTTTCACAGGTTCTCTCTTATCTTT 121
|||||

QY 631 GTGTTTTTTTCAGCCATG 650
|||||

DB 122 GTGTTTTTTTCAGCCCTG 141
|||||

RESULT 3
AL702289 463 bp mRNA linear EST 22-MAR-2002
LOCUS DKFZp686M12156_r1 686 (synonym: hlc3) Homo sapiens CDNA clone
DEFINITION DKFZp686M12156 5', mRNA sequence.
ACCESSION AL702289
VERSION AL702289.1 GI:19685644
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 463)
Bloecker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann
S. (Bloecker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and
Wiemann, S.)
Unpublished (1999)
JOURNAL
COMMENT Contact: Bloecker H
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ): Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBR (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the CDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686M12156) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 463
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686M12156"
/clone_1lb="586 (synonym: hlc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pTriplex2; Site_1: SflIA; Site_2: SflIB;
CDNA-collection"

BASE COUNT 127 a 98 c 110 g 128 t

ORIGIN

Query Match 18.2%; Score 121.8; DB 9; Length 463;
Best Local Similarity 94.7%; Pred. No. 3.2e-20; Indels 0; Gaps 0;
Matches 126; Conservative 0; Mismatches 7;

QY 486 GAGTGAATCATCTGTTGCTTTAGAGTAAGTTTACCTGCGAGTCTCTGCTG 545
|||||

DB 2 GAGTGAATCATCTGTTGCTTTAGAGTAAGTTTACCTGCGAGTCTCTGCTG 61
|||||

QY 546 GAAGTTTCTCTCTGCGAGACCAAGTTTGCCTTACGCTGAGGAGGAGTCTTTC 605
|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 20:26:44 ; Search time 1349 Seconds
(without alignments)
8043.718 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670
Sequence: 1 aagcgtcccaagacacgctt.....ggggttaataaagcgttt 670

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: EST:*
2: em_estiba:*
3: em_esthum:*
4: em_estlin:*
5: em_estinu:*
6: em_estrov:*
7: em_estrpl:*
8: em_estro:*
9: em_estl:*
10: gb_estf2:*
11: gb_hic:*
12: gb_estc3:*
13: gb_estl4:*
14: gb_estl5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_liv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	
1	151.6	22.6	9	AL600571	AL600571 DKFZ313f	
2	132	19.7	421	9	AL702116	AL702116 DKFZP686E
3	121.8	18.2	463	9	AL702289	AL702289 DKFZP686M
4	94.8	14.1	804	13	B1824550	B1824550 603035156
5	48.4	7.2	1101	17	CNS0176S	AL108022 Drosophila
6	42.4	6.3	940	14	BQ941968	BQ941968 AGENCOURT

RESULT 1	AL600571	522 bp	mRNA	linear	EST 14-AUG-2001
DERIVATION	DKFZP313F1234.F1.313 (synonym: hicc2)	Homo sapiens	cDNA clone		
ACCESSION	DKFZP313F1234.5	5'	mRNA sequence.		
VERSION	AL600571.1	GI:15164077			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Ansoorge, W., Wilkner, U., Mewes, W., Weill, B. and Wiemann, S.				
AUTHORS	EST (Ansoorge, W., Wilkner, U., Mewes, H.W., Weill, B. and Wiemann, S.)				
TITLE	Unpublished (1999)				
JOURNAL	Contact: Ansoorge W				
COMMENT					

ALIGNMENTS

Am Klopferapitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No si sequence available.
This clone (DKFZP313F1234) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

Search completed: July 10, 2003, 21:31:37
Job time : 225 secs

CC diagnosis and treatment of eye diseases, proliferative retinopathy,
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 CC Crohn's disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 6431 BP; 1843 A; 158 C; 1469 G; 2961 T; 0 other;

Query Match 5.4%; Score 36; DB 24; Length 6431;
 Best Local Similarity 47.7%; Pred. No. 3;
 Matches 105; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

CC 422 TCCCTCTTGAATGTTGGAATGGGCAATCTGCTTAAACGAAACATTCTT 481
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 98 TCGGTTTTTAAAGTGTGGCAATATAGTATAGTATAGTATAGTATAGTAT 157
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 482 GTTCGAGTGAATCATCTCTGCTTCTAGAGTAAAGTTACCTGAGTCTCTCG 541
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 158 AATGTATGTATATATATATATATATATATATATATATATATATATAT 217
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 542 TGGTGAAGTTTCTCTCTCTCGAGACCAATCTGCTTACGCTGAGGAGAGTGT 601
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 218 TTTTAT 277
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 602 TTTCACAGTTCCTCCCTTTATCTTTGTTTCTT 641
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 278 TTGTAGTTTATATTTTGTAAATATATATATATATATATATATATAT 317
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
 AAS46557
 ID AAS46557 standard; DNA; 6432 BP.

XX AAS46557;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #279.

XX Human: tumour suppressor gene; oncogene; antitumour; cytostatic;

KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SN;

KW cytosine methylation; ds.

XX Homo sapiens.

XX OS

XX WO200168912-A2.

XX PD 20-SEP-2001.

XX PF 15-MAR-2001; 2001WO-EP02955.

XX PR 15-MAR-2000; 2000DE-1013847.

XX PR 06-APR-2000; 2000DE-1019058.

XX PR 07-APR-2000; 2000DE-1019173.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI: 2001-602752/68.

XX FT Fragments of chemically modified genes associated with tumour suppressor

XX PT genes and oncogenes, useful in designing primers and probes for

XX PT analysing diseases associated with cytosine methylation state e.g.

XX PT cancer

XX PS Claim 1; SEQ ID No 279; 27pp; English.

CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pre-treated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (SS) and sequences complementary to (SS). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 6432 BP; 1845 A; 159 C; 1471 G; 2957 T; 0 other;

Query Match 5.4%; Score 36; DB 22; Length 6432;
 Best Local Similarity 47.7%; Pred. No. 3;
 Matches 105; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

CC 422 TCCCTCTTGAATGTTGGAATGGGCAATCTGCTTAAACGAAACATTCTT 481
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 98 TCGGTTTTTAAAGTGTGGCAATATAGTATAGTATAGTATAGTATAGTAT 157
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 482 GTTCGAGTGAATCATCTCTGCTTCTAGAGTAAAGTTACCTGAGTCTCTCG 541
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 158 AATGTATGTATATATATATATATATATATATATATATATATATATAT 217
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 542 TGGTGAAGTTTCTCTCTCTCGAGACCAATCTGCTTACGCTGAGGAGAGTGT 601
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 218 TTTTAT 277
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 602 TTTCACAGTTCCTCCCTTTATCTTTGTTTCTT 641
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 CC 278 TTGTAGTTTATATTTTGTAAATATATATATATATATATATATATAT 317
 CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
 ABV25186
 ID ABV25186 standard; CDNA; 2819 BP.

XX ABV25186;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker CDNA 25177.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

XX Homo sapiens.

XX OS

XX WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US05171.

XX PR 17-FEB-2000; 2000US-183319P.

XX PR 16-MAR-2000; 2000US-189862P.

XX PR 25-MAY-2000; 2000US-207454P.

XX PR 09-JUN-2000; 2000US-211314P.

XX PR 18-JUL-2000; 2000US-219007P.

CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 916 BP; 138 A; 93 C; 321 G; 364 T; 0 other;

Query Match 5.4%; Score 36.2; DB 24; Length 916;
 Best Local Similarity 50.9%; Pred. No. 1;
 Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

OY 475 ATTCTCTTTCGAGTGCATCTCTCTCTTTCGCTTTAGAGTAAGTTACCCGCACTT 534
 DB 326 ATTTCGGTTTAAATTCGGCGTTTCGTTTGAATTCGTTTCTGTTTGTGTTTATA 385
 OY 535 CCTTCGTGAGTGAAGTTCTCTCTCTCTGAGACACAGATTGCGCTTACGCTGAGG 594
 DB 386 TTTTCTTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCG 445
 OY 595 GAAGTGTTCACAGGTTCTCCTCTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCG 643
 DB 446 GAGGTTTACGCTGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCG 494

RESULT 8

ABQ49947/c
 ID ABQ49947 standard; DNA; 916 BP.

AC ABQ49947;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36538.

XX Human; cytosine methylation; 5'-CPG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

XX 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CPG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method

CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 916 BP; 364 A; 321 C; 93 G; 138 T; 0 other;

Query Match 5.4%; Score 36.2; DB 24; Length 916;
 Best Local Similarity 50.9%; Pred. No. 1;
 Matches 86; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

OY 475 ATTCTCTTTCGAGTGCATCTCTCTCTTTCGCTTTAGAGTAAGTTACCCGCACTT 534
 DB 591 ATTTCGGTTTAAATTCGGCGTTTCGTTTGAATTCGTTTCTGTTTGTGTTTATA 532
 OY 535 CCTTCGTGAGTGAAGTTCTCTCTCTCTGAGACACAGATTGCGCTTACGCTGAGG 594
 DB 531 TTTTCTTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCG 472
 OY 595 GAAGTGTTCACAGGTTCTCCTCTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCG 643
 DB 471 GAGGTTTACGCTGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCG 423

RESULT 9

ABQ67119
 ID ABQ67119 standard; DNA; 6431 BP.

AC ABQ67119;

DT 28-AUG-2002 (first entry)

DE Human angiogenesis associated polynucleotide SEQ ID NO 149.

XX Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; anti-leukers;
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;
 KW antihypertensive; antiarrhythmic; antidiabetic; antipsoriatic;
 KW antiarteriosclerotic; ds.

XX Homo sapiens.

XX WO200246454-A2.

PD 13-JUN-2002.

PF 06-DEC-2001; 2001WO-EP14320.

XX 06-DEC-2000; 2000DE-1061338.

PR (EPIC-) EPIGENOMICS AG.

PI Schacht O;

DR WPI; 2002-500450/53.

PT New nucleic acid fragments from chemically treated
 PT angiogenesis-associated genes, useful for determining methylation
 PT status, e.g. in diagnosis or treatment of cancer -

PS Claim 1; SEQ ID NO 149; 41pp + Sequence Listing; German.

XX The invention relates to a nucleic acid (I) comprising a segment of 18
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 CC having sequences (ABQ66971-ABQ67176) or their complements. (I), also
 CC related oligomers, are used to evaluate the methylation status and/or
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for

ABK84723
 ID ABK84723 standard; cDNA: 2887 BP.
 AC ABK84723;
 XX
 DT 14-AUG-2002 (first entry)
 DE Human cDNA differentially expressed in granulocytic cells #1294.
 XX
 XX Human; ss; granulocytic cell; DNA chip; bacterial infection;
 KM viral infection; parasitic infection; protozoal infection;
 KM fungal infection; sterile inflammatory disease; psoriasis;
 KM rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
 KM cardiac reperfusion injury; renal reperfusion injury; ARDS;
 KM adult respiratory distress syndrome; inflammatory bowel disease;
 KM Crohn's disease; ulcerative colitis; periodontal disease;
 KM granulocyte activation; chronic inflammation; allergy.
 XX
 OS Homo sapiens.
 XX
 PN WO200228999-A2.
 XX
 PD 11-APR-2002.
 XX
 PE 03-OCT-2001; 2001WO-US30821.
 XX
 PR 03-OCT-2000; 2000US-237189P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
 DR WPI: 2002-435328/46.
 XX
 PT Detecting granulocyte activation by detecting differential expression
 PT of genes associated with granulocyte activation, which serves as
 PT diagnostic markers that is useful for monitoring disease states and
 PT drug toxicity -
 XX
 PS Claim 1; SEQ ID NO 1294; 114pp; English.
 XX
 CC The invention relates to detecting (M1) granulocyte (GC) activation
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
 CC DNA chip analysis as given in the specification, and comparing
 CC the expression level to an expression level in an unactivated
 CC GC, where differential expression of Gs is indicative of GCA.
 CC Also included are modulating (M2) GA by contacting Gs with an agent
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)
 CC for an agent capable of modulating GCA or an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease using the
 CC gene expression profile; (3) detecting (M4) an inflammation (especially
 CC chronic) in a tissue, an allergic response in a subject, exposure of a
 CC subject to a pathogen or sterile inflammatory disease, by detecting the
 CC level of expression in a sample of the tissue of gene(s) from Gs, where
 CC the level of expression of the gene is indicative of inflammation;
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,
 CC an allergic response in a subject, exposure of a subject to a pathogen
 CC or sterile inflammatory disease, by contacting a tissue having
 CC inflammation with an agent that modulates the expression of gene(s)
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
 CC modulating Gs; M3 is useful for screening an agent capable of modulating
 CC GCA preferably in an inflammation in a tissue; M4 is useful for
 CC detecting an inflammation (especially chronic) in a tissue, an allergic
 CC response in a subject, exposure of a subject to a pathogen or sterile
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,
 CC periodontal disease, auto bacterial infection, viral infection,
 CC parasitic infection, protozoal infection, fungal infection and M5 is
 CC useful for treating one of the above conditions. The present
 CC sequence represents a gene differentially expressed in granulocytes.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pot_sequences.
 XX
 SQ Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;
 XX
 Query Match 10.7%; Score 71.6; DB 24; Length 2887;
 Best Local Similarity 94.9%; Pred. No. 2.2e-11;
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 XX
 QY 573 GATTTCGCTTACGCTGAGGGAAGTGTTCACAGGTTCTTCCTTTATCTTTGT 632
 DB 1 GATTTCGCTTCTGCTGGAGGAAGTGTTCACAGGTTCTTCCTTTATCTTTGT 60
 QY 633 GTTTTTCGAGCCATG 650
 DB 61 GTTTTTCAGCCCTG 78
 XX
 RESULT 7
 ABQ49946
 ID ABQ49946 standard; DNA: 916 BP.
 XX
 AC ABQ49946;
 XX
 DT 12-JUL-2002 (first entry)
 XX
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36537.
 XX
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KM drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 KM SNP; cell differentiation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 XX
 PE 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PR 03-SEP-2000; 2000DE-1044543.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI: 2002-371829/40.
 XX
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

/product= MORT-1 binding protein (MACH) isoform alpha

XX FT
XX PN WO9703998-A1.
XX PD 06-FEB-1997.
XX PE 14-JUN-1996; 96WO-US10521.
XX PR 16-APR-1996; 96IL-0117932.
XX PR 16-JUL-1995; 95IL-0114615.
XX PR 17-AUG-1995; 95IL-0114986.
XX PR 14-SEP-1995; 95IL-0115319.
XX PR 27-SEP-1995; 95IL-0116588.
XX PA (WEIN/) WEINURZEL H.
XX PA (YEDA) YEDA RES & DEV CO LTD.
XX PL Boidin M, Goltssev YV, Goncharov T, Wallach D;
XX DR WPI: 1997-132570/12.
XX DR P-PSDB; AAM1891.
XX PT New DNA encoding MACH protein that interacts with MORT-1 protein -
XX PT for mediating intracellular effects of FAS or TNF receptors, partic.
XX PT for regulating apoptosis in tumours, virus-infected cells etc.
XX PS Example 3: Page 116-117; 163pp; English.
XX CC AAT61396, and AAT61405-T61411 represent coding sequences for different
XX CC isoforms of MACH. MACH is a binding protein for the mediator of receptor
XX CC toxicity (MORT-1) protein. MORT-1 binds to the FAS ligand receptor
XX CC (FAS-R) death domain region, and triggers part of the cell death
XX CC signalling cascade in mammalian cells. Vectors containing these
XX CC sequences, the encoded proteins, and antibodies (Ab) against them are
XX CC used to modulate the effect of FAS-R ligand or TNF on cells that carry
XX CC FAS-R or p55-R. This is specifically for treating tumours, HIV-infected
XX CC cells or other diseased cells, by control of apoptosis/programmed cell
XX CC death. The encoded proteins are mediators of the cell death pathway
XX CC initiated by TNF and FAS-R binding, i.e. it mimics or enhances the
XX CC effect of MORT-1 where increased cytotoxicity is required. To inhibit the
XX CC effect of MORT-1, e.g. in cases of septic shock, graft rejection and
XX CC acute hepatitis, sequences encoding antisense molecules or ribozymes, or
XX CC Ab against the protein, are used. Compounds that inhibit MACH are
XX CC potentially useful for controlling MACH activity e.g. in cases of
XX CC autoimmune disease, oligodendrocyte death in multiple sclerosis or
XX CC AIDS-inhibited T-cell suicide. The MACH proteins can also be used to
XX CC isolate and characterise other proteins and receptors involved in
XX CC signalling and for Ab production. The Ab can be used to purify the new
XX CC proteins and for diagnosis of conditions involving abnormal function of
XX CC FAS-R mediated cellular effects.
XX SO Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;
XX Query Match 10.7%; Score 71.6; DB 18; Length 2887;
XX Best Local Similarity 94.9%; Pred. No. 2.2e-11;
XX Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTCGCTTTACGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 632
DB 1 GATTCGCTTTCTTCGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 60
OY 633 GTTTTTCGAGCCATG 650
DB 61 GTTTTTCAGGCCCTG 78

RESULT 5
ID AAD03915 standard; DNA; 2887 BP.
XX AAD03915;
XX 02-JUL-2001 (first entry)
DT

XX DE Human caspase-8 DNA.
XX KW Human; caspase-8; MACH; FLICE; Mc3;
XX KW cysteine protease; apoptosis; caspase expression cassette; metastasis;
XX KW tumour; cathepsin B; urokinase; proliferation; gene therapy; ds.
XX OS Homo sapiens.
XX FT Key Location/Qualifiers
XX FT CDS 292..1731
XX FT /tag= a
XX FT /product= "Human caspase-8"
XX PN WO200129232-A2.
XX PD 26-APR-2001.
XX PE 19-OCT-2000; 2000WO-US28941.
XX PR 20-OCT-1999; 99US-0160559.
XX PR 14-AUG-2000; 2000US-0225564.
XX PA (SCIO-) SCIOS INC.
XX PI Cordell B, Li Y;
XX DR WPI: 2001-290920/30.
XX DR P-PSDB; AAE00605.
XX PT Novel fusion polypeptide comprising first and second caspase subunit
XX PT separated by cleavage site not associated in nature with caspase
XX PT subunit, useful for cloning gene encoding enzymes involved in
XX PT proteolytic cleavage -
XX XX
XX Disclosure; Fig 15; 116pp; English.
XX CC The present sequence is a DNA encoding human Caspase-8 also known as
XX CC MACH, FLICE and Mc3. Caspases are a family of cysteine proteases, that
XX CC participate in the initiation and execution of apoptosis. Caspases exist
XX CC as pro-enzymes, activated by cleavage into a large and small subunit,
XX CC occurring after specific aspartic acid residues within the pro-enzyme
XX CC sequence. The present invention relates to a method for functional
XX CC cloning of genes encoding proteins or enzymes involved in proteolytic
XX CC cleavage. The invention is based on the use of caspase expression
XX CC cassettes comprising the coding sequence of a proteolytic cleavage site
XX CC flanked by sequences encoding two caspase subunits. A fusion polypeptide
XX CC comprising a first and a second caspase subunit, separated by a cleavage
XX CC site not associated in nature, is useful for cloning gene encoding
XX CC enzymes involved in proteolytic cleavage. An expression cassette
XX CC containing fusion polypeptide is used to identify a mutant cell line
XX CC deficient in an enzyme of interest and is also useful for diagnosis and
XX CC suppression of proliferation or metastases of a tumour cell characterised
XX CC by overexpression of a polypeptide (e.g. Cathepsin B or urokinase,
XX CC selectively expressed in the tumour cells). DNA encoding fusion
XX CC polypeptide is used in gene therapy.
XX SO Sequence 2887 BP; 853 A; 595 C; 625 G; 814 T; 0 other;
XX Query Match 10.7%; Score 71.6; DB 22; Length 2887;
XX Best Local Similarity 94.9%; Pred. No. 2.2e-11;
XX Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTCGCTTTACGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 632
DB 1 GATTCGCTTTCTTCGCTGGAGGAGGAGTGTTCACAGGTTCTCTTATCTTTGT 60
OY 633 GTTTTTCGAGCCATG 650
DB 61 GTTTTTCAGGCCCTG 78

RESULT 6

Seq	Sequence	45017 BP; 12919 A; 10037C; 956 G; 12495 T; 0 other:		
Query Match	77.08;	Score 516; DB 22; Length 45017;		
Best Local Similarity	94.45;	Pred. No. 7,6e-148;		
Matches	568; Conservative	0; Mismatches 30; Indels 4; Gaps 3;		
Oy	AATTGGAAGACAGGCGCCAGAGATGGAGACTCGACCCGAGACAGGGGTGATCCGGAGCAG 108			
Db	26162 AAATCGAGAGACATAGGCCAAGATGGAGATCTGACCTGACAGAGGCTGAGCC--GAGCAG 26105			
Oy	109 GCGTAAGCCAGTACGANTGAAACACAGCACACTTCCTCTCTTTTTCGAAAGATACCC 168			
Db	26104 GGGTAACCCAGTAGTGGGCTGAGCAGACACTTCCTCTCTTTT-TCGAACAATCTACCT 26046			
Oy	169 GCATTAGGCCACAGGGGCTGACTTTATCCAGTCCGGCGGGAGGAGAGAGGGCTGTCT 228			
Db	26045 ACATTGTAGCTACAGGGGCTGGCTTTTACCAGTCCGGCGGGAGGAGAGAGGGCTGTCT 25986			
Oy	229 GTGACTTCAGTGCAGAGGTTTGATCAAGAGCAAGAGGAAATCTTCATCCACAGCCTTT 288			
Db	25985 GTGACTTCAGTGCAGAGGTTTGATCAAGAGCAAGAGGAAATCTTCATCCACAGCCTTT 25922			
Oy	289 GCAGAAAGAAATGGCATATTACTTTGTCGCCCGGACAGGGGTTATTATTACTAAATGAGTCA 348			
Db	25925 GCAAGAAAGAAATGGCATATTACTTTGTCGCCCGGACAGGGGTTATTATTACTAAATGAGTCA 25866			
Oy	349 GTATAATGCTTCOCATTAAGATGCGACGGCTGGGCTTAGTTGACGCTCATCA 408			
Db	25865 GTATAATGCTTCOCATTAAGATGCGACGGCTGGGCTTAGTTGACGCTCATCA 25806			
Oy	409 ATTGTGCGCACATCCCTCTTCGATGATGTGGAAATTTGGGCATCTGTTCCTTAAACA 468			
Db	25805 ATTGTGCGCACATCCCTCTTCGATGATGTGGAAATTTGGGCATCTGTTCCTTAAACA 25744			
Oy	469 GGAACATTTCTTGTTGAGTGAATCTCTCTGTTTCTGTTAGGAGTAAATTTACCT 528			
Db	25746 GGAACATTTCTTGTTGAGTGAATCTCTCTGTTTCTGTTAGGAGTAAATTTACCT 25687			
Oy	529 GCACTTCCTCTGCGTGAAGTTTTCCTTCTCTGAGACACAGATTCGCTTTACCC 588			
Db	25686 GCACTTCCTCTGCGTGAAGTTTTCCTTCTCTGAGACACAGATTCGCTTTACCC 25622			
Oy	589 TGGAGGAGATGTTTTCACAGGTTCTCTCTCTTATATCTTTGTGTTTTTTCGAGCA 648			
Db	25626 TGGAGGAGATGTTTTCACAGGTTCTCTCTCTTATATCTTTGTGTTTTTTCGAGCCC 25567			
Oy	649 TG 650			
Db	25566 TG 25565			
RESULT 4				
AAT61405				
AAT61405 standard; cDNA; 2887 BP.				
AC	AAT61405;			
XX	29-OCT-1997 (first entry)			
DE	MACH isoform alpha coding sequence.			
KW	MACH; MORF-1 binding protein; mediator of receptor toxicity; cell death; antibody; FAS ligand receptor; FAS-R; death domain region; septic shock; tumour necrosis factor; tumour; HIV-infection; oligodendrocyte death; apoptosis/programmed cell death; p53-R; graft rejection; acute hepatitis; autoimmune disease; multiple sclerosis; AIDS-inhibited T-cell suicide; TNF; therapy; ss.			
OS	Homo sapiens.			
XX	Key			
TH	Location/Qualifiers			
CD	291..1731			
TT	/*tag= a			

Db 24906 GTGACTTCAGTGGCTGAGGTTTGATCAAGCAAGGAACCTCCATTCACAGCCCTT 24847
QY 289 GCAGAGAAAGATGGCATATTACTTGGCCGACAGAGGGTTATATATACATAATGAGTCA 348
Db 24846 GCAGAGAAAGATGGCATATTACTTGGCCGACAGAGGGTTATATATACATAATGAGTCA 24787
QY 349 GATTAATGCTTCCATATAAGCATGTCCAGCCCTGGGGCTTGTAGTTGCACGTCATGA 408
Db 24786 GTATTAATGCTTCCATATAAGCATGTCCAGCCCTGGGGCTTGTAGTTGCACGTCATGA 24727
QY 409 ATTGCTGCCACATCCCTCTTCGAATGTTGGAATGGGCATCTGTCTTAAACA 468
Db 24726 ATTGCTGCCACATCCCTCTTCGAATGTTGGAATGGGCATCTGTCTTAAACA 24668
QY 469 GGAACATTTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 528
Db 24667 GGAACATTTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 24608
QY 529 GCAATTCCTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 588
Db 24607 GCAATTCCTCTGTGAGTGAAGTATCTGTGCTTGGAGTAAAGTTACCT 24548
QY 589 TGAAGGAAGTGTTCACAGGTTCTCTCTTATCTTTGTTTTCGAGCCA 648
Db 24547 TGAAGGAAGTGTTCACAGGTTCTCTCTTATCTTTGTTTTCGAGCCA 24488
QY 649 TG 650
Db 24487 TG 24486

RESULT 3
AAK77217/c
ID AAK77217 standard; DNA: 45017 BP.
XX AAK77217:
AC AAK77217:
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32029.
XX
KW Human; immune: haematopoietic; immune/haematopoietic antigen; cancer:
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS
XX
PN MO200157182-A2.
XX
PD 09-AUG-2001.
XX
PE 17-JAN-2001; 2001MO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.

```

PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231247.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0230801.
PR 08-SEP-2000; 2000US-0230801.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0234400.
PR 14-SEP-2000; 2000US-0234401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0238935.
PR 13-OCT-2000; 2000US-0238937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 03-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX
XX Disclosure: SEQ ID NO 32028; 3071bp + sequence listing; English.
XX
XX AA54951 to AA64703 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AA62170 to AA61921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AA64703
XX to AA87694 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AA54942 to AA54550 and AA62169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 43938 BP; 12772 A; 9656 C; 9206 G; 12304 T; 0 other:

Query Match 77.0%; Score 516; DB 22; Length 43938;
Best Local Similarity 94.4%; Pred. No. 7.5e-148;
Matches 568; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 AATTGAGAGACAGGCGCAAGATGGAGATGAGACGCTGAGACGCTTGAATCCGAGCAG 108
DB 25083 AAATCGAGAGACATGGCCAAAGATGGAGATGAGACGCTGAGAGGCTGAGACC -GATCAG 25026
QY 109 GGCTAAGCCAGTACGATGATGACAGACACTTCTCTTTTTCCTTTCGAGACGATCC 168
DB 25025 GGCTAAGCCAGTACGATGATGACAGACACTTCTCTTTTTCCTTTCGAGACGATCC 24967
QY 169 GCATTTCAGCAGAGGCTGACTTACCCAGTCCGGGAGGAGAGAGAGGCTGCT 228
DB 24966 ACATTTCAGCAGAGGCTGACTTACCCAGTCCGGGAGGAGAGAGAGGCTGCT 24907
QY 229 GTGACTTCAGTGTGAGGCTTGTATCAAGGCAAGGAACTCTCTATCCACACCTTT 288

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 20:49:45 ; Search time 223 Seconds
(without alignments)
6766.094 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670
Sequence: 1 aagcgctccaagacacgatt.....ggggttaataaagcgctt 670

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.101002.*
1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	670	100.0	670	21 AAAS1792
C 2	516	77.0	43938	22 AAK77216
C 3	516	77.0	45017	22 AAK77217
4	71.6	10.7	2887	18 AAT61405
5	71.6	10.7	2887	22 AAD03915
6	71.6	10.7	2887	24 ABR84723
7	36.2	5.4	916	24 ABQ49946
C 8	36.2	5.4	916	24 ABQ49947
9	36	5.4	6431	24 ABQ67119

10	36	5.4	6432	22 AAS46557	Tumour suppressor
11	35.4	5.3	2819	23 ABV25186	Human prostate exp
C 12	35.2	5.3	301	22 AAF98635	Human ovarian canc
13	35.2	5.3	2020	14 AAO33209	Human cyclin D3 ps
14	35	5.2	256	22 AAK81668	Human immune/haema
C 15	35	5.2	6741	21 AHA10595	Gene encoding a su
16	35	5.2	19734	24 ABL33932	Human immune syste
17	34.6	5.2	581	24 ABL15812	Oligonucleotide fo
C 18	34.6	5.2	581	24 ABL15813	Oligonucleotide fo
19	34.2	5.1	13573	24 ABL33869	Human immune syste
C 20	34	5.1	631	24 ABL25138	Oligonucleotide fo
21	34	5.1	631	24 ABL25139	Oligonucleotide fo
22	34	5.1	5318	24 ABL54318	Chemically treated
23	34	5.1	5518	24 ABL80019	Human chemically m
24	34	5.1	5518	24 ABL32217	Human immune syste
25	34	5.1	7148	22 AAS45303	Chemically pretrea
26	34	5.1	7148	24 ABL80035	Human chemically m
27	34	5.1	7148	24 ABL28142	DNA transcription
28	34	5.1	13249	24 ABL70131	Chemically treated
29	34	5.1	13249	24 ABL32116	Human immune syste
30	34	5.1	13249	24 ABL31176	Signal transductio
31	33.8	5.0	5860	24 ABL70523	Chemically treated
32	33.8	5.0	5860	24 AAS61372	Human gene regulat
33	33.8	5.0	5860	24 ABL31456	Signal transductio
34	33.8	5.0	16891	24 ABL80058	Human chemically m
35	33.8	5.0	16891	24 ABL32652	Human immune syste
36	33.8	5.0	16891	24 ABL34504	Human metastasis a
C 37	33.6	5.0	1992	21 AAC33992	Arabidopsis thalia
38	33.6	5.0	5371	24 ABL34295	Human immune syste
39	33.6	5.0	6104	24 ABL33124	Human immune syste
40	33.6	5.0	6577	22 AAS46718	Tumour suppressor
41	33.2	5.0	2022	13 ABL31877	Cyclin D3 pseudoge
C 42	33.2	5.0	26747	24 ABL67784	Nucleotide sequenc
43	33	4.9	6833	24 ABL33145	Human immune syste
44	33	4.9	8897	23 ABL20304	Drosophila melanog
45	33	4.9	16167	24 ABL70254	Chemically treated

ALIGNMENTS

RESULT 1
AAA51792 standard; DNA: 670 BP.
AAA51792:
AC AAA51792;
XX
DT 31-OCT-2000 (first entry)
XX
DE CASP8 promoter Region 1.
XX
KW CASP8; caspase-8; promoter; Region 1; inactivation; methylation;
KW Cysteine protease; tumour suppressor; chromosome 2q33-34; neuroblastoma;
KW cancer; death receptor; apoptosis; cytosolic; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN W0200039347-A1.
XX
PD 06-JUL-2000.
XX
PF 30-DEC-1999; 99WO-US31280.
XX
PR 31-DEC-1998; 98US-0114308.
XX
PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Kidd VJ, Lahti JM, Teitz T;
XX
DR WPT; 2000-45242/39.
XX
PT Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or
prognosing cancer, comprises detecting a modification of genomic DNA


```
/note="L1MA10 repeat: matches 5479. .6315 of consensus"
repeat_region L1MA10 repeat: matches 5479. .6315 of consensus"
12979. .13263
/note="Alusx repeat: matches 9. .290 of consensus"
Alusx repeat: matches 9. .290 of consensus"
13548. .13778
/note="MIR repeat: matches 15. .262 of consensus"
MIR repeat: matches 15. .262 of consensus"
14035. .14188
/note="MER5A repeat: matches 43. .189 of consensus"
MER5A repeat: matches 43. .189 of consensus"
14785. .14859
/note="L1ME3A repeat: matches 5785. .5859 of consensus"
L1ME3A repeat: matches 5785. .5859 of consensus"
15050. .15303
/note="L1ME3A repeat: matches 5525. .5777 of consensus"
L1ME3A repeat: matches 5525. .5777 of consensus"
15317. .15774
/note="L1M4 repeat: matches 3951. .4454 of consensus"
L1M4 repeat: matches 3951. .4454 of consensus"
15337. .15774
/note="L1 repeat: matches 3951. .4433 of consensus"
L1 repeat: matches 3951. .4433 of consensus"
15943. .16022
/note="L2 repeat: matches 2664. .2750 of consensus"
L2 repeat: matches 2664. .2750 of consensus"
/note="L2 repeat: matches 2664. .2750 of consensus"
complement(16038. .16616)
/note="match: GSS: Em:AQ541664"
17138. .17193
/note="MIR repeat: matches 78. .133 of consensus"
MIR repeat: matches 78. .133 of consensus"
18611. .18904
/note="MIR repeat: matches 193. .466 of consensus"
MIR repeat: matches 193. .466 of consensus"
18905. .19230
/note="Alusg repeat: matches 1. .310 of consensus"
Alusg repeat: matches 1. .310 of consensus"
19231. .19419
/note="MIR repeat: matches 1. .193 of consensus"
MIR repeat: matches 1. .193 of consensus"
join(19542. .19799,20379. .20477)
/gene="dJ380B8.3"
join(<19542. .19799,20379. .>20477)
/gene="dJ380B8.3"
/product="dJ380B8.3 (a novel mRNA)"
/note="match: ESTs: Em:R79825"
/evidence=not_experimental
19669. .20180
/gene="dJ380B8.3"
/note="match: GSS: Em:AQ401011"
20155. .20184
/note="15 copies 2 mer ac 100% conserved"
20482. .20806
/note="LTR16C repeat: matches 56. .387 of consensus"
LTR16C repeat: matches 56. .387 of consensus"
21088. .21177
/note="FLAM A repeat: matches 10. .99 of consensus"
FLAM A repeat: matches 10. .99 of consensus"
21463. .22143
/note="HERVL repeat: matches 4345. .5045 of consensus"
HERVL repeat: matches 4345. .5045 of consensus"
21469. .21856
/note="HERVL repeat: matches 4644. .5045 of consensus"
HERVL repeat: matches 4644. .5045 of consensus"
22192. .23403
/note="HERVL repeat: matches 2953. .4189 of consensus"
HERVL repeat: matches 2953. .4189 of consensus"
23429. .24388
/note="HERVL repeat: matches 1902. .2891 of consensus"
HERVL repeat: matches 1902. .2891 of consensus"
23611. .24388
/note="HERVL repeat: matches 1902. .2701 of consensus"
HERVL repeat: matches 1902. .2701 of consensus"
24949. .25225
/note="HERV16 repeat: matches 882. .1164 of consensus"
HERV16 repeat: matches 882. .1164 of consensus"
25786. .26094
/note="Alu repeat: matches 1. .307 of consensus"
Alu repeat: matches 1. .307 of consensus"
26354. .26477
```

```
repeat_region /note="L1ME repeat: matches 458. .585 of consensus"
26716. .26792
/note="L2 repeat: matches 2666. .2741 of consensus"
L2 repeat: matches 2666. .2741 of consensus"
misc_feature complement(26932. .27500)
27047. .27174
/note="match: GSS: Em:AQ417969"
repeat_region /note="MIR repeat: matches 65. .196 of consensus"
MIR repeat: matches 65. .196 of consensus"
27177. .27238
/note="L2 repeat: matches 2644. .2709 of consensus"
L2 repeat: matches 2644. .2709 of consensus"
27243. .27405
repeat_region

Query Match 6.3% Score 42.2; DB 9; Length 166882;
Best Local Similarity 59.7%; Pred. No. 0.12; 48; Indels 0; Gaps 0;
Matches 71; Conservative 0; Mismatches 0;

QY 328 TATTATTACTAATGAGTCAGTATTAATGCTTCCATTAAGCATGTCAGCGCTGGG 387
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163927 TATTATTTCATATGTAATAATAAAGACCTTATTAAGACCTTACCTTTTGGT 163986

QY 388 CTTAGTTTCAGCTCAGCAATGCTCTGCGACATCCCTCTGTAATGTTGAATG 446
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163987 ACTAAGTTTCAAGACCAAGTATGTTTACACTTCGCGCATATGCGACTTTGAAATAG 164045
```

Search completed: July 11, 2003, 04:06:58
Job time : 2080 secs

FEATURES

Location/Qualifiers
1..223387
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="RP23-10B21"
/clone_id="RPC1 mouse BAC library 23"

BASE COUNT 66248 a 45169 c 45091 g 65176 t 1703 others
ORIGIN

Query Match 6.4%; Score 42.8; DB 2; Length 223387;
Best Local Similarity 51.6%; Pred. No. 0.078;
Matches 98; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 452 CTCGTCCTTAAACGAGAACATTTCTGTGCGATGATCTCTCTGCTTA 511
DB 84132 CTCCTCTCCAGACAGCCACCCATATATGCTCTGCGACCATTTAGCCAC 84073
QY 512 GGACTAAAGTTACCTCGAGTTCCTCTGTGAGTTTCTTTCTCGAGAGC 571
DB 84072 GGACAACTCAATCCTTGTGCTACTTCACCCCTCACCCCTCACTGAGCTAACCC 84013
QY 572 AGATTCGCTTACGCTGAGAGAGTGTTCACAGTTCCTCTTTATCTTTG 631
DB 84012 AGCCTCAGTTTCCCTTGCAGAGAGTGTGAGAGCCTTTCTTTCTTTT 83953
QY 632 TGTCTTTT 641
DB 83952 TTTTCTTTT 83943

RESULT 15

AL136307 186882 bp DNA linear PRI 01-AUG-2000
LOCUS AL136307
DEFINITION Human DNA sequence from clone RP3-380B8 on chromosome 6p24.1-25.3
Contains a gene encoding the protein Neuritin, which is involved in
proliferation of neurite outgrowth, a private kinase (PKM2)
pseudogene, a novel mRNA, 4 Cpg islands, ESTs, STS and GSSs,
complete sequence.

ACCESSION AL136307
VERSION AL136307.12 GI:8574124
KEYWORDS HMG: CPG Island; Neuritin; PKM2.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 186882)
AUTHORS Brown, J.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humgen@sanger.ac.uk
requests: clonerequests@sanger.ac.uk

COMMENT

On Jun 20, 2000 this sequence version replaced gi:8573783.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, SW, SWISSPROT, Tr, TREMBL, Wp, WORMPEP, Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6

FEATURES

source

RP3-380B8 is from the library RPC1-3 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP3-380B8 the true left
end of clone RP11-525021 is at 141303 in this sequence. The true
right end of clone RP1-182016 is at 57138 in this sequence.

Location/Qualifiers
1..186882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p24.1-25.3"
/clone_lib="RP3-380B8"
/clone_id="RPC1-3"
247..465
/note="MIR repeat: matches 16..236 of consensus
MIR repeat: matches 16..236 of consensus"
1442..2391
/note="LIMB3 repeat: matches 5281..6174 of consensus
LIMB3 repeat: matches 5281..6174 of consensus"
2435..2499
/note="MIR repeat: matches 77..145 of consensus
MIR repeat: matches 77..145 of consensus"
2540..3786
/note="LIP2 repeat: matches 4900..6146 of consensus"
2542..3786
/note="LIP5 repeat: matches 4900..6143 of consensus"
4034..4653
/note="LIM4 repeat: matches 4022..4632 of consensus
LIM4 repeat: matches 4022..4632 of consensus"
4654..4953
/note="AluX repeat: matches 1..300 of consensus
AluX repeat: matches 1..300 of consensus"
4954..6496
/note="LIM9 repeat: matches 4632..6308 of consensus
LIM9 repeat: matches 4632..6308 of consensus"
6522..6820
/note="MLTID repeat: matches 14..283 of consensus
MLTID repeat: matches 14..283 of consensus"
6821..6890
/note="MER67A repeat: matches 374..543 of consensus
MER67A repeat: matches 374..543 of consensus"
6987..7184
/note="MLTID repeat: matches 308..505 of consensus
MLTID repeat: matches 308..505 of consensus"
7290..7462
/note="MIR repeat: matches 71..262 of consensus"
7357..7462
/note="MIR repeat: matches 71..177 of consensus"
8007..8093
/note="LIM4 repeat: matches 5688..5772 of consensus
LIM4 repeat: matches 5688..5772 of consensus"
8094..8329
/note="LIM8 repeat: matches 6029..6173 of consensus
LIM8 repeat: matches 6029..6173 of consensus"
8932..8994
/note="L2 repeat: matches 2637..2699 of consensus"
8996..9199
/note="MER33 repeat: matches -9..192 of consensus
MER33 repeat: matches -9..192 of consensus"
9200..9559
/note="THE1B repeat: matches 1..364 of consensus
THE1C repeat: matches 1..371 of consensus"
9560..9685
/note="MER33 repeat: matches 192..324 of consensus
MER33 repeat: matches 192..324 of consensus"
10364..10771
/note="MLTIG repeat: matches 65..512 of consensus
MLTIG repeat: matches 65..512 of consensus"
10775..10875
/note="MIR repeat: matches 134..252 of consensus"
12145..12963

```

repeat_region /note="AluY repeat: matches 1. .308 of consensus"
24237. .24500
/note="AluY repeat: matches 1. .289 of consensus"
repeat_region 24534. .24824
/note="AluSC repeat: matches 1. .293 of consensus"
repeat_region 25038. .25173
/note="MIR repeat: matches 105. .246 of consensus"
repeat_region 25739. .26024
/note="AluY repeat: matches 1. .303 of consensus"
repeat_region 26120. .26431
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 27285. .27401
/note="FLAM_C repeat: matches 1. .117 of consensus"
misc_feature 27846. .28076
/note="match: STR: Em:G24508"
repeat_region 29035. .29346
/note="AluS5 repeat: matches 1. .307 of consensus"
misc_feature complement(29346. .29399)
/note="match: GSS: Em:AQ469605"
misc_feature complement(29346. .29501)
/note="match: STR: Em:HSP42F07"
misc_feature complement(join(29371. .29596,32193. .32262))
/note="match: GSS: Em:AQ475563"
misc_feature complement(29371. .29598)
/note="match: GSS: Em:AQ349959"
misc_feature complement(join(29378. .29596,32193. .32398))
/note="match: GSS: Em:AQ677478"
repeat_region 29724. .30076
/note="LIMC/D repeat: matches 5540. .5906 of consensus"
repeat_region 30159. .30678
/note="LIMC4 repeat: matches 6758. .7354 of consensus"
repeat_region 30792. .31104
/note="LIMC repeat: matches 1794. .2123 of consensus"
repeat_region 31204. .31564

```

Query Match 6.4%: Score 42.8; DB 9; Length 138006;
 Best Local Similarity 45.7%: Pred. No. 0.073;
 Matches 149; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

```

QY 316 GCGCAGACGGGTTATTACTAATGAGTCAATGATTAATGCTTCCATTAAGCATGT 375
DB 48019 GGTACAGTATTTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 48078
QY 376 CCAGCGTCGGGCTTGTGTCAGTCGTCATGATTTGTCGACATCCCTCTTGAAT 435
DB 48079 TGTGTTTGTCAACATATGTTTAACTGCTGATGAACGCTCTTAACTTTTGTTCATT 48138
QY 436 GGTGGAATGGGATCTGTCCTTTAAAGAGAAACATTTCTTGTGAGTGAATCA 495
DB 48139 TTTTAAATAGAGTCTGCTGTTTCTTATTGAGTGTGAATGTTCTATGATTTTGA 48198
QY 496 TCTGTGTCGCTTAGAGTAAGTTTACCTGTCAGTCTCTCTGCTGGTGAAGTTTCT 555
DB 48199 TATCGGTTCTTTTGAATACATGTTTGAACAAATATTTTCTCATGCTGTGCTGCG 48258
QY 556 CTTCCTCGGAGACCGATTCGCTTACGCGTGGAGGAAGTTTTCACAGGTTCTG 615
DB 48259 TTTTCACTTTTGTAGTGCATATTTCAAGTACAGAAAGTTCTTAATTTAATGAGTCA 48318
QY 616 CTCCTTTTATCTTTGTGTTTGT 641
DB 48319 TTAATTTTGTGTTTGTGTTTGT 48344

```

RESULT 14
 AC079470/c 223387 bp DNA linear HTG 02-SEP-2000
 LOCUS Mus musculus clone RP23-10B21, WORKING DRAFT SEQUENCE, 18 ordered
 DEFINITION pieces.
 AC079470
 AC079470.1 GI:9964835
 VERSION HTG; HTGS_PHASE2; HTGS_DRAFT.
 KEYWORDS Mus musculus.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 223387)
 DOE Joint Genome Institute.
 Sequencing of Mouse
 Unpublished
 2 (bases 1 to 223387)
 DOE Joint Genome Institute.
 Direct Submission
 Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

 Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: <http://www.jgi.doe.gov>

Project Information
 Center Project Name: 1740757
 Center clone name: RP23-10B21

Summary Statistics

Consensus quality: 212195 bases at least Q40
 Consensus quality: 219485 bases at least Q30
 Consensus quality: 220848 bases at least Q20
 Estimated insert size: 208000; agarose-fp estimation
 Estimated insert size: 222587; sum-of-contigs estimation
 Quality coverage: 10.31 in Q20 bases; agarose-fp estimation
 Quality coverage: 9.63 in Q20 bases; sum-of-contigs estimation.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 18 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 been provided by the submitter.

This sequence will be replaced
 by the finished sequence as soon as it is available and
 the accession number will be preserved.

```

* 1 24944: contig of 24944 bp in length
* 24945 25044: gap of unknown length
* 25045 31043: contig of 5999 bp in length
* 31044 31143: gap of unknown length
* 31144 49115: contig of 17972 bp in length
* 49116 49215: gap of unknown length
* 49216 67918: contig of 18703 bp in length
* 67919 68018: gap of unknown length
* 68019 79367: contig of 11349 bp in length
* 79368 79468: gap of unknown length
* 79469 85787: contig of 6320 bp in length
* 85788 85888: gap of unknown length
* 85889 92600: contig of 6713 bp in length
* 92601 92700: gap of unknown length
* 92701 95687: gap of 2987 bp in length
* 95688 95787: gap of unknown length
* 95788 100230: contig of 4443 bp in length
* 100231 100330: gap of unknown length
* 100331 116173: contig of 15843 bp in length
* 116174 116273: gap of unknown length
* 116274 121323: contig of 5050 bp in length
* 121324 121423: gap of unknown length
* 121424 124686: contig of 3263 bp in length
* 124687 124786: gap of unknown length
* 124787 154386: contig of 29600 bp in length
* 154387 154486: gap of unknown length
* 154487 164212: contig of 9726 bp in length
* 164213 164312: gap of unknown length
* 164313 169021: contig of 4709 bp in length
* 169022 169121: gap of unknown length
* 169122 171322: contig of 2200 bp in length
* 171323 211231: gap of unknown length
* 211232 211331: contig of 39810 bp in length
* 211332 223387: contig of 12056 bp in length.

```

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL, Sw, SWISSPROT, Tr, TrEMBL, Wp, WormPep, information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormPep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RPl1-294H11 is from the library RPl1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm> VECTOR: PBACE3.6

IMPORTANT: This sequence is not the entire insert of clone RPl1-294H11. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPl3-448K1 is at 137907 in this sequence. The true right end of clone RPl1-600I9 is at 100 in this sequence.

FEATURES

source

```
1. 138006
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RPl1-294H11"
/clone_lib="RPl1-11.2"
1. 100
/note="match: STS: Em:HS60019B"
2014. 2059
/note="23 copies 2 mer tg 97% conserved"
2149. 2446
/note="MTR1B repeat: matches 96. 390 of consensus"
2466. 2583
/note="AluYb repeat: matches 1. 118 of consensus"
4339. 4637
/note="AluY repeat: matches 1. 298 of consensus"
4842. 5151
/note="AluSP repeat: matches 1. 312 of consensus"
5152. 5332
/note="AluSg repeat: matches 134. 312 of consensus"
5803. 6023
/note="AluYb repeat: matches 77. 289 of consensus"
6056. 6364
/note="AluY repeat: matches 1. 306 of consensus"
/note="complement(6200. 6688)
/note="match: GSS: Em:AQ528743"
complement(6225. 6677)
/note="match: GSS: Em:AQ524727"
complement(6414. 6688)
/note="match: GSS: Em:AQ564844"
6902. 7096
/note="MER20 repeat: matches 1. 218 of consensus"
7660. 7756
/note="HAL1 repeat: matches 60. 161 of consensus"
7883. 8555
/note="LIMC4 repeat: matches 6798. 7502 of consensus"
8556. 8860
/note="AluSP repeat: matches 1. 308 of consensus"
8861. 8904
/note="LIMC4 repeat: matches 7502. 7543 of consensus"
complement(8906. 9376)
/note="match: GSS: Em:AQ479008"
9175. 9469
/note="match: GSS: Em:AQ434542"
```

```
repeat_region 9221. 9246
/note="13 copies 2 mer tg 96% conserved"
repeat_region 9258. 9466
/note="HERV1 repeat: matches 2068. 2279 of consensus"
repeat_region 9482. 9677
/note="MER45 repeat: matches 1. 178 of consensus"
repeat_region 9681. 9901
/note="HERV16 repeat: matches 1794. 2013 of consensus"
repeat_region 9902. 10208
/note="AluSx repeat: matches 1. 310 of consensus"
repeat_region 10209. 10314
/note="HERV16 repeat: matches 1689. 1794 of consensus"
repeat_region 10348. 10586
/note="HERV16 repeat: matches 1357. 1577 of consensus"
repeat_region 10621. 10915
/note="AluY repeat: matches 1. 295 of consensus"
repeat_region 10952. 10991
/note="MER52A repeat: matches 1063. 1107 of consensus"
repeat_region 11073. 11271
/note="HERV16 repeat: matches 981. 1197 of consensus"
repeat_region 11411. 11987
/note="HERV16 repeat: matches 140. 813 of consensus"
repeat_region 11988. 12295
/note="AluSx repeat: matches 4. 303 of consensus"
repeat_region 12296. 12444
/note="HERV16 repeat: matches 2. 140 of consensus"
repeat_region 12753. 13081
/note="AluY repeat: matches 1. 308 of consensus"
repeat_region 13088. 13206
/note="MIR repeat: matches 102. 223 of consensus"
repeat_region 13335. 13611
/note="AluSx repeat: matches 29. 305 of consensus"
repeat_region 13930. 14527
/note="LIM4A repeat: matches 5706. 6298 of consensus"
repeat_region 15521. 15805
/note="AluSx repeat: matches 11. 296 of consensus"
repeat_region 16061. 16130
/note="charlies repeat: matches 32. 108 of consensus"
repeat_region 16922. 17050
/note="FLM1C repeat: matches 1. 143 of consensus"
repeat_region 17052. 17237
/note="AluSg/x repeat: matches 104. 274 of consensus"
repeat_region 17806. 18234
/note="LOR1B repeat: matches 32. 460 of consensus"
repeat_region 18338. 18414
/note="MTR11 repeat: matches 329. 402 of consensus"
repeat_region 18415. 18720
/note="AluYb repeat: matches 1. 304 of consensus"
repeat_region 18721. 18966
/note="MTR11 repeat: matches 68. 329 of consensus"
repeat_region 19033. 19350
/note="AluSx repeat: matches 1. 301 of consensus"
repeat_region 19791. 19897
/note="LIMB6 repeat: matches 5096. 5205 of consensus"
repeat_region 19904. 20370
/note="LIMB6 repeat: matches 5159. 5621 of consensus"
repeat_region 20372. 20760
/note="MSTRC repeat: matches 1. 401 of consensus"
repeat_region 20764. 21227
/note="LIMB6 repeat: matches 5625. 6100 of consensus"
repeat_region 21235. 21530
/note="AluSg repeat: matches 1. 296 of consensus"
repeat_region 21600. 21713
/note="MER93 repeat: matches 19. 136 of consensus"
repeat_region 21714. 22019
/note="AluSg repeat: matches 1. 305 of consensus"
repeat_region 22020. 22219
/note="MER93 repeat: matches 136. 394 of consensus"
repeat_region 22220. 22316
/note="MER57B repeat: matches 296. 394 of consensus"
misc_feature 22950. 23214
/note="match: GSS: Em:B85271"
repeat_region 23820. 24103
```

```

* 4342 4441: gap of unknown length
* 4442 6030: contig of 1589 bp in length
* 6031 6130: gap of unknown length
* 6131 7908: contig of 1778 bp in length
* 7909 8008: gap of unknown length
* 8009 9035: contig of 1027 bp in length
* 9036 9135: gap of unknown length
* 9136 10139: contig of 1004 bp in length
* 10140 10239: gap of unknown length
* 10240 11425: contig of 1186 bp in length
* 11426 11525: gap of unknown length
* 11526 13078: contig of 1553 bp in length
* 13079 13178: gap of unknown length
* 13179 14359: contig of 1181 bp in length
* 14360 14459: gap of unknown length
* 14460 15625: contig of 1166 bp in length
* 15626 15725: gap of unknown length
* 15726 17134: contig of 1409 bp in length
* 17135 17234: gap of unknown length
* 17235 18728: contig of 1494 bp in length
* 18729 18828: gap of unknown length
* 18829 20366: contig of 1538 bp in length
* 20367 20466: gap of unknown length
* 20467 21995: contig of 1529 bp in length
* 21996 22095: gap of unknown length
* 22096 23292: contig of 1197 bp in length
* 23293 23392: gap of unknown length
* 23393 24954: contig of 1562 bp in length
* 24955 25054: gap of unknown length
* 25055 26072: contig of 1018 bp in length
* 26073 26172: gap of unknown length
* 26173 27383: contig of 1211 bp in length
* 27384 27483: gap of unknown length
* 27484 28663: contig of 1180 bp in length
* 28664 28763: gap of unknown length
* 28764 29866: contig of 1103 bp in length
* 29867 31660: gap of unknown length
* 31661 31760: contig of 1694 bp in length
* 31761 32761: gap of unknown length
* 32762 32861: contig of 1001 bp in length
* 32862 34306: gap of unknown length
* 34307 34406: gap of unknown length
* 34407 36015: contig of 1609 bp in length
* 36016 36115: gap of unknown length
* 36116 37733: contig of 1618 bp in length
* 37734 37833: gap of unknown length
* 37834 39472: contig of 1639 bp in length
* 39473 39572: gap of unknown length
* 39573 41586: contig of 2014 bp in length
* 41587 41686: gap of unknown length
* 41687 43005: contig of 1319 bp in length
* 43006 43105: gap of unknown length
* 43106 44919: contig of 1814 bp in length
* 44920 45019: gap of unknown length
* 45020 46210: contig of 1191 bp in length
* 46211 46310: gap of unknown length
* 46311 48629: contig of 2319 bp in length
* 48630 48729: gap of unknown length
* 48730 50329: contig of 1600 bp in length
* 50330 50429: gap of unknown length
* 50430 51800: contig of 1371 bp in length
* 51801 51900: gap of unknown length
* 51901 54894: contig of 2994 bp in length
* 54895 54994: gap of unknown length
* 54995 57436: contig of 2442 bp in length
* 57437 57536: gap of unknown length
* 57537 59281: contig of 1745 bp in length
* 59282 59381: gap of unknown length
* 59382 61694: contig of 2313 bp in length
* 61695 61794: gap of unknown length
* 61795 64543: contig of 2749 bp in length
* 64544 64643: gap of unknown length

```

```

* 64644 66047: contig of 1404 bp in length
* 66048 66147: gap of unknown length
* 66148 68418: contig of 2271 bp in length
* 68419 68518: gap of unknown length
* 68519 70306: contig of 1788 bp in length
* 70307 72336: gap of unknown length
* 72337 72437: contig of 1930 bp in length
* 72438 74495: gap of unknown length
* 74496 74595: contig of 2059 bp in length
* 74596 76864: gap of unknown length
* 76865 76964: contig of 2269 bp in length
* 76965 78379: gap of unknown length
* 78380 78480: contig of 1415 bp in length
* 78481 80190: gap of unknown length
* 80191 80290: contig of 1711 bp in length
* 80291 83933: gap of unknown length
* 83934 84033: contig of 3643 bp in length
* 84034 85931: gap of unknown length
* 85932 86031: contig of 1898 bp in length
* 86032 86031: gap of unknown length
* 86032 86087: contig of 2056 bp in length
* 86088 88187: gap of unknown length
* 88188 90268: contig of 2081 bp in length
* 90269 90369: gap of unknown length
* 90369 92824: contig of 2456 bp in length
* 92825 92825: gap of unknown length
* 92925 95808: contig of 2884 bp in length
* 95809 95908: gap of unknown length
* 95909 98559: contig of 2651 bp in length

Query Match
Best Local Similarity 50.0%; Score 43.2; DB 2; Length 184700;
Matches 108; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 426 TCTTGTGATGGTTGGAATGGGACATCTCTGCTTAAACAGAAACATTCTTGTTC 485
DB 81232 TCCCTACTAGCTGATTTAGATGCTTACCTCAAGCCTGTTCTGGACTTGT 81291

QY 486 GAGTAGACATCTGCTCTGCTTTAGAGAAAGTTAACCTCAGTCTCTGCGGT 545
DB 81292 CAGACATCTCAGTCTGTTGCTCTGCTGGAGAGGAGGATCCCAATCTCT 81351

QY 546 GAGATTCTCTCTCTCTCGGAGACAGATTCTGCTTTAAGCTGGAGGAGATTTC 605
DB 81352 GCTCTTGTCTGCTGCTCTGCTGAGGACACATCTCTAATAGTCTAATGAGGATC 81411

QY 606 AAGGTTCTCTCTCTTTATCTTTGTTT 641
DB 81412 ACAGTCTCCACATAGTTTGTGTTTGT 81447

RESULT 13
AL390074 138006 bp DNA linear PRI 05-APR-2001
LOCUS Human DNA sequence from clone RP11-294H11 on chromosome 6, complete
DEFINITION
ACCESSION AL390074
VERSION AL390074.17 GI:13560018
KEYWORDS HTG.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 138006)
REFERENCE Tracey, A.
Direct Submission
Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgehire,
CB10 1SA, UK. E-mail enquiries: humquer@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Apr 6, 2001 this sequence version replaced GI:13396608.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the

```


	Query Match	10.7%	Score 71.6;	DB 6;	Length 2887;	
	Best Local Similarity	94.9%	Pred. No.2.9e-11;			
	Matches 74; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;	
Qy	573 GATTCGCTCCATTACGCTGGAGGAAGTGTTCACAGGTTCTCCTCTTAATCTTTGT	632				
	1 GATTCGCTTTCTCGCTGGAGGAGGAGTTTTCACAGGTTCTCCTCTTAATCTTTGT	60				
Qy	633 GTTTTTTTTGAGGCATG	650				
Db	61 GTTTTTTTTCAAGCCCTG	78				
	RESULT 9					
LOCUS	AX134419	2887 bp	DNA	linear	PAT 29-MAY-2001	
DEFINITION	Sequence 15 from Patent WO0129232.					
VERSION	AX134419					
KEYWORDS	AX134419.1 GI:14271065					
SOURCE	.					
ORGANISM	human.					
	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
REFERENCE	1 (bases 1 to 2887)					
AUTHORS	Cordell,B. and Li,Y.					
TITLE	Functional cloning of genes encoding proteins/enzymes involved in proteolytic cleavage					
JOURNAL	Patent: NO 0129232-A 15 26-Apr-2001;					
	Sciis Inc. (US)					
FEATURES	Location/Qualifiers					
source	1..2887					
	/organism="Homo sapiens"					
CDS	/db_xref="taxon:9606"					
	292..1731					
	/note="unnamed protein product"					
	/codon_start=1					
	/protein_id="CAC39526.1"					
	/db_xref="GI:14271066"					
	/translation="MDPSRMLDYGEDLSDSEDLASLKFLSLDYIPORKEPIKDALML FORLOERMLEESNLSFLKELFIRINDLLITLRKEEMERELOTGPRAQISAVR VMLOISEEVSSELRSFKLEIOEKSKLDODMNLIDIFIEKERVILGEGLDIILL KRVAOINKSLKLTIIINEERSKRRSSLESPPESGECILGVMTIISPSPRODESE SQTDADYQMKSRPGRCYLTIINNPNFPARPKAPVKILSIIPRNQTHIDAGALITTFEE LHEIKRPDDCTVEQIYEILKITYQLDMHSNDCTICCLISBGDCGITYIGDGQAAPYIL ELTSQETGLKCPSLAGRKFVFIQACGDNYQKGIPEVTDSERPYLEMDISSQTR IPDADPFLGMATVNNCVSYRNPAGEFTWIOSLOSRLRECRPDGDIILTITEVNEHV SNKDDKKRMGROMPPTFLTKKLVPFS"					
BASE COUNT	853 a 595 c 625 g 814 t					
ORIGIN						
	Query Match	10.7%;	Score 71.6;	DB 6;	Length 2887;	
	Best Local Similarity	94.9%;	Pred. No.2.9e-11;			
	Matches 74; Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;	
Qy	573 GATTCGCTTACCTGGAGGAAGTGTTCACAGGTTCTCCTCTTAATCTTTGT	632				
	1 GATTCGCTTTCTCGCTGGAGGAGGAGTTTTCACAGGTTCTCCTCTTAATCTTTGT	60				
Qy	633 GTTTTTTTTGAGGCATG	650				
Db	61 GTTTTTTTTCAAGCCCTG	78				
	RESULT 10					
LOCUS	HSNACHA1	2887 bp	mRNA	linear	PRI 25-SEP-1997	
DEFINITION	H.sapiens mRNA for MACH-alpha-1 protein.					
ACCESSION	X98172					
VERSION	X98172.1 GI:1403318					
KEYWORDS	cell death; MACH-alpha-1; protease.					
SOURCE	Homo sapiens.					
ORGANISM	Homo sapiens					

REFERENCE	JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS		Mammalia; Ethidea; Primates; Catarrhini; Homidae; Homo.
TITLE		1 (bases 1 to 2887)
REFERENCE	JOURNAL	Boldin,M.P., Goncharov,T.M., Goltsev,Y.V. and Wallach,D.
AUTHORS		Involvement of MACH, a novel MORF1/FAD-dependent protease, in
TITLE		Fas/Apo-1- and TNF receptor-induced cell death
PUBMED		Cell 85 (6), 803-815 (1996)
MEDLINE		96279826
REFERENCE		8681376
AUTHORS		2 (bases 1 to 2887)
TITLE		Wallach,D.
REMARK		Direct Submission
AUTHORS		Submitted (24-MAY-1996) D. Wallach, The Weizmann Institute, Dept of
JOURNAL		Membrane Research & Biophysics, Rehovot 76100, ISRAEL
TITLE		Revised by [3]
REFERENCE		3 (bases 1 to 2887)
AUTHORS		Wallach,D.
TITLE		Direct Submission
JOURNAL		Submitted (17-JUN-1996) D. Wallach, The Weizmann Institute, Dept of
REFERENCE		Membrane Research & Biophysics, Rehovot 76100, ISRAEL
AUTHORS		4 (bases 1 to 2887)
TITLE		Mandruzzato,S., Brasseur,F., Andry,G., Boon,T. and van der
FEATURES		Bruggen,P.
CDS		A CASP-8 mutation recognized by cytolytic T lymphocytes on a human
JOURNAL		head and neck carcinoma
MEDLINE		J. Exp. Med. 186 (5), 785-793 (1997)
PUBMED		97419196
FEATURES		9271594
SOURCE		Location/Qualifiers
		1..2887
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/tissue-type="thymus"
		292..1731
		/function="protease"
		/note="triggers cell death"
		/codon_start=1
		/product="MACH-alpha-1"
		/protein_id="CA68853.1"
		/db_xref="gi:2440071"
		/db_xref="swiss:PROT:Q14790"
		/translation="MDFSRLNYDIGEQLDSEDLASLKLFLSLDYIPORKOEIKDALMD
		FORLOEKMLEESNLSPFKLLPRINRDLITVLTNRKEMERLOTGGRAGLSAVY
		VMLQIISPEERSSELRSFKLLOEISCKADDNNLDITFIMKKRYILGEGLDI
		KRVQAOTIKSLKIINDYEERSKERSSLESPPENSGEGLGVMTSDSPRODS
		SQTLDKYOKSKRPKRGYCLINNHNFAKAKVKRIHSIRRNCHHDAGAATTFEE
		LHEIFKPHDDCTVEOITELIKLYQMBSNMDCFTCCILSGDGIGIIHTGGCAPTI
		ELTSQETGLKCPSLAGRKVFIAQCQDNQKGIPIVTDEEDPYLEMDUSPQTR
		IIPDADELPLGMATVNVCYSYNPAEGTYIOSLOSCLRERCPRDDITLITLVENEY
		SNKDKRMKGROMPOPTFLTKRLVFPSD"
BASE COUNT		853 a 595 c 625 g 814 t
ORIGIN		
Query Match		10.7%: Score 71.6; DB 9; Length 2887;
Best Local Similarity		94.9%: Pred. No. 2.9e-11;
Matches		74; Conservative 0; MisMatches 4; Indels 0; Gaps 0;
OY	573 GATTGTGCCTTAGCGCTGGAGGAAGCTTTTCACAGGCTTCCTCCTTAATCTTTGT	632
DB	1 GATTGTGCCTTCTCTCTGGAGGAAAGTGTTTTCACAGGTTCTCTCTTTATCTTTGT	60
OY	633 GTTTTTTTCGAGCATG 650	
DB	61 GTTTTTTTCAGCCCTG 78	
RESULT 11		
LOCUS	166494	7218 bp DNA linear PAT 28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.	
ACCESSION	166494	
VERSION	166494.1 GI:2724471	
KEYWORDS		

BASE COUNT	853 a	595 c	625 g	814 t	ORIGIN
Query Match	10.7%	Score 71.6	DB 9	Length 2887	
Best Local Similarity	94.9%	Pred No. 2.9e-11			
Matches 74	Conservative 0	Mismatches 4	Indels 0	Gaps 0	
QY	573	GATTCGCGCTTTACGCTGAGAGGAGTGTTCACAGGTTCTCCTTTATCTTTGT	632		
Db	1	GATTCGCGCTTTCTCTGTGAGAGGAGTGTTCACAGGTTCTCCTTTATCTTTGT	60		
QY	633	GTTTTTTTCGAGCATG	650		
Db	61	GTTTTTTTCAGCCCTG	78		
RESULT 11					
I66494					
LOCUS	I66494		7218 bp	DNA	linear
DEFINITION	Sequence 14 from patent US 5670367.				
ACCESSION	I66494				
VERSION	I66494.1	GI:2724471			
KEYWORDS					

QY 561 TCTCGAGAGACAGATCTGCTTACGCTGGAGGAGAGTGTTCACAGTTCTCTC 618
 Db 61 TCTCGAGAGACAGATCTGCTTACGCTGGAGGAGAGTGTTCACAGTTATATTC 118

RESULT 6
 AF422929
 LOCUS
 DEFINITION Homo sapiens clone 4 procaspase-8 (CASP8) mRNA, partial cds;
 ACCESSION AF422929
 VERSION AF422929.1 GI:19401529
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.
 TITLE The procaspase-8 isoform, procaspase-8L, recruited to the BAP31 complex at the endoplasmic reticulum
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (77), 4331-4336 (2002)
 MEDLINE 21927603
 PUBMED 11917123

REFERENCE
 2 (bases 1 to 777)
 Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-2001) Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada

FEATURES
 source
 1. 777
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="4"
 1..>777
 /gene="CASP8"
 266..>777
 /gene="CASP8"
 /note="caspase precursor; alternatively spliced"
 /codon_start=1
 /product="procaspase-8"
 /protein_id="AAU87632.1"
 /db_xref="GI:19401530"
 /translation="MDFSRNLYDIGEOLDESDIASLKLSDYIPQKOEPIKDALMT
 FORLEKRMLEESNLSFKELFLRINLDLITYLNTREEMERELQTPGKQISAYR
 VMLYQISEEVSRESLSFKFLQELRISCKLDDMDLDFIEMERVILGSKLDIL
 KRYCAQINSL"

BASE COUNT 218 a 156 c 203 g 194 t

Query Match 10.7%; Score 71.6; DB 9; Length 777;
 Best Local Similarity 94.9%; Pred. No. 2.5e-11;
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTGCGCTTACGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 632
 Db 1 GATTGCGCTTACGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 60

QY 633 GTTTTTCGAGCCATG 650
 Db 61 GTTTTTCGAGCCCTG 78

RESULT 7
 AF422925
 LOCUS
 DEFINITION Homo sapiens procaspase-8L (CASP8) mRNA, complete cds;
 ACCESSION AF422925
 VERSION AF422925.1 GI:19401518
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.
 TITLE The procaspase-8 isoform, procaspase-8L, recruited to the BAP31 complex at the endoplasmic reticulum
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (77), 4331-4336 (2002)
 MEDLINE 21927603
 PUBMED 11917123

REFERENCE
 2 (bases 1 to 2503)
 Breckenridge,D.G., Nguyen,M., Kuppig,S., Reth,M. and Shore,G.C.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-2001) Biochemistry, McGill University, 3655 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada

FEATURES
 source
 1. 2503
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..2503
 /gene="CASP8"
 115..1731
 /gene="CASP8"
 /note="caspase precursor; alternatively spliced"
 /codon_start=1
 /product="procaspase-8L"
 /protein_id="AAU87628.1"
 /db_xref="GI:19401519"
 /translation="MEGGRARVYTESRNFLLGAPTPPAPAEVLEGRIGDSEAMY
 FKGAGDILLPFKMDSRNLYDIGEOLDESDIASLKLSDYIPQKOEPIKDALM
 LEORLEKRMLEESNLSFKELFLRINLDLITYLNTREEMERELQTPGKQISAY
 RMLYQISEEVSRESLSFKFLQELRISCKLDDMDLDFIEMERVILGSKLDIL
 LKRYCAQINSLKIINDYEFSEKERSLSGDEFSNGELGVMITSDSPEDQS
 ESOTLDKRYOKRSKRGYCLIIHNHNFARAEKXPKLSHISDRNGTLDGALTYTE
 ELHPIKRPDCTVQEQIYELIKIYIOMDSMDCDFICITLSHGKGLIYGDGEARI
 YELNSQFTGLCKPSLAKGRKRYFTQACQGDGVQGIPTVETDSERQPIEMDSRPR
 YIPDADFLGMATVNNCVSYRNPAEGTWIQSQCQSLREKCPRGDDILITILEVNE
 VSNKDKRNMKGMPQPTFLRKRLVPSD"

BASE COUNT 726 a 528 c 593 g 656 t

Query Match 10.7%; Score 71.6; DB 9; Length 2503;
 Best Local Similarity 94.9%; Pred. No. 2.8e-11;
 Matches 74; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 573 GATTGCGCTTACGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 632
 Db 1 GATTGCGCTTACGCTGGAGGAGAGTGTTCACAGTTCCTCTTATCTTTGT 60

QY 633 GTTTTTCGAGCCATG 650
 Db 61 GTTTTTCGAGCCCTG 78

RESULT 8
 AR211527
 LOCUS
 DEFINITION Sequence 14 from patent US 6399327.
 ACCESSION AR211527
 VERSION AR211527.1 GI:21514872
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 2887)
 Wallach,D., Boldin,M., Goncharov,T. and Golsteyn,Y.V.
 TITLE Modulators of the function of FAS receptors and other proteins
 JOURNAL Patent: US 6399327-A 14 04-JUN-2002;
 FEATURES
 source
 1..2887
 /organism="unknown"
 BASE COUNT 853 a 595 c 625 g 814 t

QY 349 GTATAATGCTTTCCATAAAGCATGTCACGCGCTGCGGCTTTAGTTGACGTCATGA 408
 DB 159575 GTATAAAGCTTTCCATAAAGCATGTCACGCGCTTGTAGTTGACGTCATGA 159516
 QY 409 ATTGCTGCGACATCCCTCTCTGTAATGGATGGGATGCGATCTGCTCTTTAAACA 468
 DB 159515 ATTGCTGCGACATCCCTCTCTGTAATGGATGGGATGCGATCTGCTCTTTAAACA 159457
 QY 469 GGAACATTTCTTTGTCAGTGTACATCTGTTCTCTTTAGAGATTAAGTTTACCT 528
 DB 159456 GGAACATTTCTTTGTCAGTGTACATCTGTTCTCTTTAGAGATTAAGTTTACCT 159397
 QY 529 GCAGTCTCTTCTGTGTGAAGTTTCTTTCTCTCGAGACACATCTCTTACGC 588
 DB 159396 GCAGTCTCTTCTGTGTGAAGTTTCTTTCTCTCGAGACACATCTCTTACGC 159337
 QY 589 TGGAGGGAAGTGTTCACAGTCTCTCTCTTTATCTTTTGTGTTTTCAGACCA 648
 DB 159336 TGGAGGGAAGTGTTCACAGTCTCTCTCTTTATCTTTTGTGTTTTCAGACCC 159277
 QY 649 TG 650
 DB 159276 TG 159275

RESULT 4
 AF210257 982 bp DNA linear PRI 06-JUN-2000
 LOCUS Homo sapiens caspase 8 gene, 5' flanking region.
 DEFINITION AF210257
 ACCESSION AF210257.1 GI:8272474
 VERSION
 KEYWORDS
 ORGANISM Homo sapiens.
 SOURCE Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 982)
 Teitz,T., Wei,T., Valentine,M.B., Vanin,E.F., Grenet,J.,
 Valentine,V.A., Behm,F.G., Look,A.T., Lahti,J.M. and Kidd,V.J.
 Caspase 8 is deleted or silenced preferentially in childhood
 neuroblastomas with amplification of MYCN
 Nat. Med. 6 (5), 529-535 (2000)
 JOURNAL
 MEDLINE 10802708
 PUBMED 20264561
 TITLES 2 (bases 1 to 982)
 Teitz,T., Wei,T., Valentine,M.B., Vanin,E.F., Grenet,J.,
 Valentine,V.A., Behm,F.G., Look,A.T., Lahti,J.M. and Kidd,V.J.
 Direct Submission
 Submitted (01-DEC-1999) Tumor Cell Biology, St. Jude Children's
 Research Hospital, 332 N. Lauderdale, Memphis, TN 38105, USA
 JOURNAL
 TITLE Location/Qualifiers
 FEATURES
 source
 1. 982
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2q33-q34"
 misc_feature
 1. 982
 /note="5' flanking region of caspase 8 gene"
 BASE COUNT 198 a 235 c 241 g 308 t
 ORIGIN

Query Match 54.6%; Score 366; DB 9; Length 982;
 Best Local Similarity 97.2%; Pred. No. 1.6e-105;
 Matches 383; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 257 GCAAGGGAACCTCTCTATTCAGACCTTTGCAAGAAATGCGATATCTGCGG 316
 DB 1 GCAAGGGAACCTCTCTATTCAGACCTTTGCAAGAAATGCGATATCTGCGCA 60
 QY 317 CCACAGGGGTATTTACTAATGAGTCAATTAATGCTTTCAATTAAGCATGTC 376
 DB 61 CCACAGGGGTATTTACTAATGAGTCAATTAATGCTTTCAATTAAGCATGTC 120

QY 377 CAGCGCTCGGCGCTTTAGTTGACAGTCATGATTTGTCTGCCACATCCCTTCTGATG 436
 DB 121 CAGCGCTCGGCGCTTTAGTTGACAGTCATGATTTGTCTGCCACATCCCTTCTGATG 180
 QY 437 GTTGAATTTGGGATCTCTCTCTCTTTAAACAGAAACATTTCTTTGAGTGTAT 496
 DB 181 GTTGAATTTGGGATCTCTCTCTCTTTAAACAGAAACATTTCTTTGAGTGTAT 239
 QY 497 CTCTGTTCTCTTTAGGAGATTAAGTTTACCTGACAGTCTCTTGTGTGAATTTCTTC 556
 DB 240 CTCTGTTCTCTTTAGGAGATTAAGTTTACCTGACAGTCTCTTGTGTGAATTTCTTC 299
 QY 557 TTCTCGGAGAGACCATTTCTGCTTTAGCGTGGAGGGAAGTTTTCACAGTTCTTC 616
 DB 300 TTCTCGGAGAGACCATTTCTGCTTTAGCGTGGAGGGAAGTTTTCACAGTTCTTC 359
 QY 617 TCCTTTATCTTTTGTGTTTTCAGACCATG 650
 DB 360 TCCTTTATCTTTTGTGTTTTCAGACCATG 393

RESULT 5
 AF422926 413 bp mRNA linear PRI 03-APR-2002
 LOCUS Homo sapiens clone 1 procaspase-8 (CASP8) mRNA, partial cds;
 DEFINITION AF422926
 ACCESSION AF422926.1 GI:19401520
 VERSION
 KEYWORDS
 ORGANISM Homo sapiens.
 SOURCE Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 413)
 Breckenridge,D.G., Nguyen,M., Kuppis,S., Reith,M. and Shore,G.C.
 The procaspase-8 isoform, procaspase-8L, recruited to the Bap1
 complex at the endoplasmic reticulum
 Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4331-4336 (2002)
 JOURNAL
 MEDLINE 21927603
 PUBMED 11917123
 TITLES 2 (bases 1 to 413)
 Breckenridge,D.G., Nguyen,M., Kuppis,S., Reith,M. and Shore,G.C.
 Direct Submission
 Submitted (24-SEP-2001) Biochemistry, McGill University, 3655
 Promenade Sir William Osler, Montreal, PQ H3G1Y6, Canada
 JOURNAL
 TITLE Location/Qualifiers
 FEATURES
 source
 1. 413
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="1"
 1. >413
 /gene="CASP8"
 136..>413
 /note="CASP8"
 /note="caspase precursor: alternatively spliced"
 /codon_start=1
 /product="procaspase-8"
 /protein_id="AAL87629.1"
 /db_xref="GI:19401521"
 /translation="MDPSRLYDIDGEOLESDIASIKFLSDYIPKROEPIKALML
 FQROEKMLEBSNLSFKELLPRIKLDLITYLNTKREMERELQTP"
 BASE COUNT 115 a 87 c 96 g 115 t
 ORIGIN

Query Match 16.2%; Score 108.4; DB 9; Length 413;
 Best Local Similarity 94.9%; Pred. No. 3.9e-23;
 Matches 112; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 501 GTTCGCTTTAGAGTAAGTTTACCTGCACTCTCTGTGTGAAGTTTCTCTTC 560
 DB 1 GTTCGCTTTAGAGTAAGTTTACCTGCACTCTCTGTGTGAAGTTTCTCTTC 60

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenoe, M., Catanesi, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-156B7; the clone sequenced to the right is RP11-53618, 200 bp overlap. Actual start of this position is at base position 1 of RP11-575C6; actual end is at base position 180956 of RP11-575C6.

The sequence from base position 134614 to 134714 is derived from PCR product of RP11-575C6 BAC DNA.

FEATURES
The clone RP11-575C6 contains a transposon in the vector.

Source Location/Qualifiers

```

1..181150
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-575C6"
/clone_lib="RPCI-11"
1..354
/rpt_family="L1"
91..95
/note="similar to Homo sapiens EST BG495251 (NID:g13456786)"
repeat_region
124..153
/rpt_family="(TC)n"
repeat_region
192..228
/rpt_family="(T)n"
repeat_region
330..1055
/rpt_family="L1"
553..620
/rpt_family="T-rich"
repeat_region
1056..1363
/rpt_family="Alu"
repeat_region
1210..1243
/rpt_family="(T)n"
repeat_region
1228..1246
/note="match to EST AW444732 (NID:g6986494)"
misc_feature
1364..2264
/rpt_family="L1"
repeat_region
3338..3550
/rpt_family="MIR"
repeat_region
3739..3866
/rpt_family="MIR"
misc_feature
3971..4133
/note="similar to Homo sapiens EST AU117593 (NID:g10932573)"
misc_feature
4041..4495
/note="similar to Homo sapiens EST AI459220 (NID:g4391202) to08f03.x1"
misc_feature
4053..4639
/note="similar to Mus musculus EST BF163999 (NID:g1044285)"
misc_feature
4092..4830
/note="similar to Homo sapiens EST AL528010 (NID:g12791503)"
misc_feature
4128..4340
/note="match to EST AV760501 (NID:g10918349)"
misc_feature
4184..4742
```

```

/note="match to EST AI052250 (NID:g3308241) o21c07.x1"
misc_feature
4339..4903
/note="similar to Homo sapiens EST AU117593 (NID:g10932573)"
misc_feature
4485..4491
/note="similar to Homo sapiens EST BE788113 (NID:g10209311)"
misc_feature
4657..5468
/note="similar to Homo sapiens EST AU563032 (NID:g12912044)"
misc_feature
4712..5400
/note="match to EST AU131936 (NID:g10992290)"
misc_feature
4780..5582
/note="similar to Homo sapiens EST BG164297 (NID:g12671000)"
misc_feature
4963..5682
/note="similar to Homo sapiens EST AU124988 (NID:g10948704)"
misc_feature
5106..5633
/note="similar to Mus musculus EST A116325 (NID:g3516649) ue97c04.y1"
misc_feature
5169..5675
/note="similar to Homo sapiens EST AU144381 (NID:g11005902)"
misc_feature
5194..5761
/note="similar to Homo sapiens EST BE529656 (NID:g11617019)"
misc_feature
5237..6019
/note="similar to Homo sapiens EST AU130508 (NID:g10990862)"
misc_feature
5238..6043
/note="similar to Mus musculus EST BF166846 (NID:g11047198)"
misc_feature
5326..5653
/note="similar to Homo sapiens EST BG390376 (NID:g13283824)"
misc_feature
5361..6051
/note="similar to Homo sapiens EST AU135897 (NID:g10956436)"
misc_feature
5361..5812
/note="similar to EST BG862638 (NID:g14213176)"
misc_feature
5420..6048
/note="similar to Homo sapiens EST BE788113 (NID:g10209311)"
misc_feature
5551..6327
/note="similar to Homo sapiens EST AU130771 (NID:g10991125)"
misc_feature
5615..5984
/note="similar to Homo sapiens EST AA280659 (NID:g1923454)

Query Match      77.0%: Score 516; DB 9; Length 181150;
Best Local Similarity 94.4%: Pred. No. 2.9e-153;
Matches 568; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY      49 AATTGAGAGACAGGGCCAGAGATGGAGACAGCGTGGATCCGAGAGAG 108
DB      159872 AACTGAGAGACATGGCCAGAGATGGAGACATGGATCCGAGAGAG 159815
QY      105 GCGTAAGCCAACTAGCAATGAACAGACCACTCTCTTTTCTGTAAGACTTACCC 168
DB      159814 GCGTAAGCCAACTAGCGGTGAGACCACTCTCTCTTTTCTGTAAGACTTACCT 159756
QY      165 GCATTTCAGCCACAGGGCTGACTTTACCCAGTCCGCGGAGGAGAGAGAGAGGCTGCT 228
DB      159755 ACATTTCAGCTACAGGCTGCTTACCCAGTCCGCGGAGGAGAGAGAGGCTGCT 159696
QY      229 GTACTTCAGTCTAGCTAGTTTGATCAAGGCAAGGGAAGTCCATTCACAGACCCCTT 288
DB      159695 GTGACTTCAGTCTAGCTAGTTTGATCAAGGCAAGGGAAGTCCATTCACAGACCCCTT 159636
QY      288 GCAAGAAAGATGCGATATTAAGTTCGCGCCGACAGAGGGGTATATTAATTAATGAGTCA 348
DB      159635 GCAAGAAAGATGCGATATTAAGTTCGCGCCGACAGAGGGGTATATTAATTAATGAGTCA 159576
```


TITLE
Rouleau, G.A., Ikeda, J.-F. and Hayden, M.R.
Cloning and characterization of three novel genes, ALS2CL1, ALS2CL2, and ALS2CL3, in the juvenile amyotrophic lateral sclerosis (ALS2) critical region at chromosome 2q33-q34: candidate genes for ALS2

JOURNAL
Genomics 71 (2), 200-213 (2001)

MEDLINE
21100893

REFERENCE
2 (bases 1 to 6502)

AUTHORS
Hadano, S., Ikeda, J., and Hayden, M.R.

JOURNAL
Submitted (24-FEB-2000) Shinji Hadano, Tokai University, The Institute of Medical Sciences; Bonseida, Isehara, Kanagawa 259-1193, Japan (E-mail: shinji@ngs.med.u-tokai.ac.jp, Tel: +81-463-91-5095, Fax: +81-463-91-4993)

FEATURES
Location/Qualifiers
1..6502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q33"
1..1478
/note="Intron is approximately 24.0 kbp long"
/number=2
1479..1743
/product="caspase-8"
/note="CDS is reported in Acc#: AB038985
alternative 5'UTR"
/number=3
1744..3924
/number=3
3925..3970
/product="caspase-8"
/note="CDS is reported in Acc#: AB038985
alternative 5'UTR"
/number=4
3971..56502
/note="Intron is approximately 5.9 kbp long"
Intron
1718 a 1361 c 1504 g 1919 t
BASE COUNT
ORIGIN

Query Match 77.0%; Score 516; DB 9; Length 6502;
Best Local Similarity 94.4%; Pred. No. 2e-153;
Matches 568; Conservative 0; Mismatches 30; Indels 4; Gaps 3;

QY 49 AATTGGAAGACAGGSCCAAGATGGAGCTGAGCTGAGCAGCGGTTGATCCGAGCAG 108
Db 959 AACTGGAGACACTGGCCCAAGATGGAGCTGAGCTGAGCAGCGGTTGATCCGAGCAG 1016

QY 109 GGCTTAAGCCAGTACGAATGAACCAAGCACTTCTCTTTTCTGACGATCTACCC 168
Db 1017 GGCTTAAGCCAGTACGAATGAACCAAGCACTTCTCTTTTCTGACGATCTACCC 1075

QY 169 GCATTTCAGCAGAGGCTGACTTACCGAGTCCGGCGGAGGAGGAGGAGGCTGCT 228
Db 1076 ACATTTCAGCAGAGGCTGACTTACCGAGTCCGGCGGAGGAGGAGGAGGCTGCT 1135

QY 229 GTGACTCAGTGTGAGGTTTGAATCAAGGCAAGGAACTTCTTCCAGACCTTT 288
Db 1136 GTGACTCAGTGTGAGGTTTGAATCAAGGCAAGGAACTTCTTCCAGACCTTT 1195

QY 289 GCAGAGAAAGATGGCATATTACTTGGCGGACAGGAGGTTATTACTAATGAGTCA 348
Db 1196 GCAGAGAAAGATGGCATATTACTTGGCGGACAGGAGGTTATTACTAATGAGTCA 1255

QY 349 GATTAATGCTTCCATTAAGCATCCAGCGCTGGGCTTATGTTTCAGCTCCATGA 408
Db 1256 GATTAATGCTTCCATTAAGCATCCAGCGCTGGGCTTATGTTTCAGCTCCATGA 1315

QY 409 ATTGTCTGCACATCCCTCTTCTGATGTTGTTGAATTTGGCATCTGTGTTCTTTAA 468
Db 1316 ATTGTCTGCACATCCCTCTTCTGATGTTGTTGAATTTGGCATCTGTGTTCTTTAA 1374

QY 469 GGAACATTCTTCTGAGTGAATCACTCTGTTGTTAGGATTAAGTTACCT 528
Db 1375 GGAACATTCTTCTGAGTGAATCACTCTGTTGTTAGGATTAAGTTACCT 1434

QY 529 GCAGTCTCTTGTGTAAGTTTCTTCTCTGAGACAGCATTCGCTTACGC 568
Db 1435 GCAGTCTCTTGTGTAAGTTTCTTCTCTGAGACAGCATTCGCTTACGC 1494

QY 569 TGGAGGAGAGTCTTTCACAGGTTCTCTCTTATCTTGTGTTTTCGAGCCA 648
Db 1495 TGGAGGAGAGTCTTTCACAGGTTCTCTCTTATCTTGTGTTTTCGAGCCC 1554

QY 649 TG 650
Db 1555 TG 1556

RESULT 2
AC074016
LOCUS
DEFINITION
Homo sapiens chromosome 2 clone RP11-155P18, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
AC074016 179216 bp DNA linear HTG 09-AUG-2001
AC074016
AC074016.5 GI:15144340
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 179216)
Waterston, R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 179216)
Waterston, R.H.
Direct Submission
Submitted (10-JUL-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA
On Aug 9, 2001 this sequence version replaced gi:1956173.

COMMENT
Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0155P18
Summary Statistics
Sequencing vector: pL3; 48k
Sequencing vector: pL3; 48k
Chemistry: Dye-terminator Big Dye; 44% of reads
Chemistry: Dye-terminator Big Dye; 44% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178615 bases at least Q40
Consensus quality: 178729 bases at least Q20
Consensus quality: 178811 bases at least Q20
Insert size: 174000; agarose-fp
Insert size: 181802; sum-of-contigs
Quality coverage: 15.53 in Q20 bases; agarose-fp
Quality coverage: 14.71 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 82838: contig of 82838 bp in length
* 82839 82838: gap of unknown length
* 82939 179216: contig of 96278 bp in length.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2003, 22:21:06 ; Search time 1995 Seconds

(without alignments)
9773.875 Million cell updates/sec

Title: US-09-477-082-1

Perfect score: 670
Sequence: 1 aaagcctcacaagacgatt.....ggggttaataaagcgttt 670

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pin.*
35: em.htg.pod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	516	77.0	6502	9 AB03898052	AB038981 Homo sapi
2	516	77.0	179216	2 AC074016	AC074016 Homo sapi
3	516	77.0	181150	9 AC0007256	AC0007256 Homo sapi
4	366	54.6	982	9 AF210257	AF210257 Homo sapi
5	108.4	16.2	413	9 AF422926	AF422926 Homo sapi
6	71.6	10.7	777	9 AF422929	AF422929 Homo sapi
7	71.6	10.7	2503	9 AF422925	AF422925 Homo sapi
8	71.6	10.7	2887	6 AR211527	AR211527 Sequence
9	71.6	10.7	2887	6 AX134419	AX134419 Sequence
10	71.6	10.7	2887	9 HSMACHA1	X98172 H.sapiens m
11	49.6	7.4	7218	6 I66494	I66494 Sequence 14
12	43.2	6.4	184700	2 AC112822	AC112822 Rattus no
13	42.8	6.4	138006	2 AL330074	AL330074 Human DNA
14	42.8	6.4	223387	2 AC079470	AC079470 Mus muscu
15	42.2	6.3	186882	9 AL136307	AL136307 Human DNA
16	41.8	6.2	148334	2 AC087106	AC087106 Homo sapi
17	40.6	6.1	49430	2 AC100434	AC100434 Mus muscu
18	40	6.0	121538	9 AC128707	AC128707 Homo sapi
19	40	6.0	145002	2 AC018372	AC018372 Homo sapi
20	40	6.0	148208	2 AC010951	AC010951 Homo sapi
21	40	6.0	160325	2 AC073044	AC073044 Homo sapi
22	40	6.0	176887	2 AC041032	AC041032 Homo sapi
23	40	5.8	186581	2 AL506810	AL506810 Homo sapi
24	38.8	5.8	209518	2 AC098348	AC098348 Rattus no
25	38.6	5.8	36037	9 AC003966	AC003966 Homo sapi
26	38.6	5.8	157443	9 AC109352	AC109352 Homo sapi
27	38.6	5.8	158804	2 AC123019	AC123019 Rattus no
28	38.4	5.7	141400	9 AP004581	AP004581 Homo sapi
29	38.2	5.7	137227	2 HUM97612	AF075004 Homo sapi
30	38.2	5.7	69303	9 AL357496	AL357496 Human DNA
31	38.2	5.7	72045	9 AC092214	AC092214 Homo sapi
32	38.2	5.7	164018	2 AC074242	AC074242 Homo sapi
33	38.2	5.7	17387	2 AC104753	AC104753 Mus muscu
34	38.2	5.7	209685	2 AC026231	AC026231 Mus muscu
35	38.2	5.7	114409	2 AL161742	AL161742 Human DNA
36	38	5.7	137227	2 AL356914	AL356914 Homo sapi
37	38	5.7	158622	2 AC117302	AC117302 Rattus no
38	38	5.7	160175	2 AC090178	AC090178 Homo sapi
39	38	5.7	163502	9 AC023456	AC023456 Homo sapi
40	38	5.7	17386	9 AC016343	AC016343 Homo sapi
41	38	5.7	182107	9 AC015798	AC015798 Homo sapi
42	38	5.6	110042	9 AL590788	AL590788 Human DNA
43	37.8	5.6	146098	9 AC074112	AC074112 Homo sapi
44	37.8	5.6	182431	2 AC068380	AC068380 Homo sapi
45	37.8	5.6	182431	2 AC068380	AC068380 Homo sapi

ALIGNMENTS

RESULT 1
AB03898052
LOCUS AB03898052 6502 bp DNA linear PRI 23-FEB-2001
DEFINITION Homo sapiens gene for caspase-8, exon 3, exon 4.
ACCESSION AB038981
VERSION AB038981.1 GI:12862688
KEYWORDS
SEGMENT
SOURCE
ORGANISM Homo sapiens DNA.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Hadano, S., Yanagisawa, Y., Skaug, J., Fichter, K., Nasir, J.,
Martindale, D., Koop, B.F., Scherer, S.W., Nicholson, D.W.,

Search completed: July 8, 2003, 06:04:42
Job time : 25.6593 secs

```

: Sequence 868, Application US/09969373
: Patent No. US2002013852A1
: GENERAL INFORMATION:
: APPLICANT: Efiertz, Roger J.
: APPLICANT: Haugue, Brian M.
: TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
: FILE REFERENCE: 38-10(52679)/A
: CURRENT APPLICATION NUMBER: US/09/969,373
: CURRENT FILING DATE: 2001-10-02
: PRIOR APPLICATION NUMBER: US 09/754,853
: PRIOR FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 09/760,427
: PRIOR FILING DATE: 2001-01-13
: PRIOR APPLICATION NUMBER: US 09/855,768
: PRIOR FILING DATE: 2001-05-15
: NUMBER OF SEQ ID NOS: 4593
: SEQ ID NO 868
: LENGTH: 132
: TYPE: DNA
: ORGANISM: Glycine max
US-09-969-373-868

```

```

Query Match      72.8%; Score 18.2; DB 10; Length 132;
Best Local Similarity 87.0%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY      3 ATATATCTACATTCACAAACA 25
      ||||||| ||||| |||
DB      53 ATATATATATATATTCACAAATCA 31

```

```

RESULT 15
US-10-066-543-2507
: Sequence 2507, Application US/10066543
: Publication No. US20030087818A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Yugu
: APPLICANT: Pyle, Ruth A.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Indrias, Carol Joseph
: APPLICANT: Lodes, Michael J.
: APPLICANT: Secrist, Heather
: APPLICANT: Carter, Darrick
: APPLICANT: Fanger, Gary R.
: APPLICANT: Smith, Carole L.
: APPLICANT: Durham, Margarita
: APPLICANT: Stolk, John A.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.563
: CURRENT APPLICATION NUMBER: US/10/066,543
: CURRENT FILING DATE: 2002-01-31
: NUMBER OF SEQ ID NOS: 3417
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2507
: LENGTH: 234
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: LOCATION: 115, 127, 156, 163, 208
: OTHER INFORMATION: n = A,T,C or G
US-10-066-543-2507

```

```

Query Match      72.8%; Score 18.2; DB 9; Length 234;
Best Local Similarity 87.0%; Pred. No. 7.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY      3 ATATATCTACATTCACAAACA 25
      ||||||| ||||| |||
DB      83 ATATATATATACATTGAAAAA 105

```

1 TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
2 TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
3 TITLE OF INVENTION: THERAPY OF BREAST CANCER
4 FILE REFERENCE: MRI-049
5 CURRENT APPLICATION NUMBER: US/10/198,846
6 CURRENT FILING DATE: 2002-07-18
7 PRIOR APPLICATION NUMBER: 60/306,220
8 PRIOR FILING DATE: 2001-07-18
9 NUMBER OF SEQ ID NOS: 14084
10 SOFTWARE: FASTSEQ for Windows Version 4.0
11 SEQ ID NO 13758
12 LENGTH: 4020
13 TYPE: DNA
14 ORGANISM: Homo sapiens
15 FEATURE:
16 NAME/KEY: misc_feature
17 LOCATION: 1, 2, 1631, 4018, 4019, 4020
18 OTHER INFORMATION: n = A,T,C or G
19 US-10-198-846-13758

Query Match 74.4%; Score 18.6; DB 9; Length 4020;
Best Local Similarity 84.0%; Pred. No. 7.5e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25
1713 CCATATATATATATATCCAGACAA 1737

RESULT 11
US-10-239-676-78/C
Sequence 78, Application US/10239676
Publication No. US20030082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT APPLICATION NUMBER: US/10/239,676
CURRENT FILING DATE: 2002-09-24
PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10018058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1

PRIOR FILING DATE: 2001-04-06
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO 78
LENGTH: 6071
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (2618, 2780)
US-10-239-676-78

Query Match 74.4%; Score 18.6; DB 9; Length 6071;
Best Local Similarity 84.0%; Pred. No. 7.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25
3475 CAAATAATATTACATTCATAAACA 3451

RESULT 12

US-09-754-853A-2
Sequence 2, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parnell, Laurence D.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
CURRENT FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174,880
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 2
LENGTH: 335913
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: (45163)..(45314)..(45450)..(45509)..(46941)..(48763)..(48975)..(49573)
OTHER INFORMATION: Clone ID: 240017_region_g3
US-09-754-853A-2

Query Match 74.4%; Score 18.6; DB 9; Length 335913;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25
20113 CGATATATATATACATAAACA 201137

RESULT 13
US-09-754-853A-3
Sequence 3, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parnell, Laurence D.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
CURRENT FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174,880
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 3
LENGTH: 335913
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: (46798)..(48763)..(48975)..(49573)
OTHER INFORMATION: Clone ID: 240017_region_g3
US-09-754-853A-3

Query Match 74.4%; Score 18.6; DB 9; Length 335913;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATTCATAAACA 25
20113 CGATATATATATACATAAACA 201137

RESULT 14
US-09-969-373-868/C

US-09-918-995-9724

Query Match 74.4% Score 18.6; DB 9; Length 469;
Best Local Similarity 84.0%; Pred. No. 5.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAGACAA 25
|||||
DB 130 CCATATATATATATATCCAGACAA 154

RESULT 7
US-09-918-995-13475

; Sequence 13475, Application US/09918995
; Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756 FROM VARIOUS CDNA LIBRARIES

CURRENT APPLICATION NUMBER: US/09/918,995

PRIOR FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 13475

LENGTH: 487

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc.feature

LOCATION: (1)...(487)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-13475

Query Match 74.4% Score 18.6; DB 9; Length 487;

Best Local Similarity 84.0%; Pred. No. 5.7e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAGACAA 25
|||||
DB 355 CCATATATATATATATCCAGACAA 379

RESULT 8
US-09-745-763-12

; Sequence 12, Application US/09745763

; Patent No. US20020065394A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

McCoy, John M.

LaValle, Edward R.

Collins-Racie, Lisa A.

Evans, Cheryl

Merberg, David

Trecay, Maurice

Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

ENCODING THEM

NUMBER OF SEQUENCES: 219

CORRESPONDENCE ADDRESS:

ADDRESS: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: MA

COUNTRY: U.S.A.

ZIP: 02140

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/745,763

FILING DATE: 18-Jun-2000

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sprunger, Suzanne A.

REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8284

TELEFAX: (617) 876-5851

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1463 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 12:

US-09-745-763-12

Query Match 74.4% Score 18.6; DB 10; Length 1463;

Best Local Similarity 84.0%; Pred. No. 6.6e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAGACAA 25
|||||
DB 1167 CCATATATATATATATCCAGACAA 1191

RESULT 9
US-10-044-205A-21

; Sequence 21, Application US/10044205A

; Patent No. US20020123464A1

GENERAL INFORMATION:

APPLICANT: BANBARU, Rajasekhar

TITLE OF INVENTION: 69087, 15821, and 15418, Methods and Compositions of Human Pro

FILE REFERENCE: 10147-5201

CURRENT APPLICATION NUMBER: US/10/044,205A

CURRENT FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: US 60/242,428

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: US 60/241,884

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: US 60/241,877

NUMBER OF SEQ ID NOS: 44

SOFTWARE: Patentin version 3.1

SEQ ID NO 21

LENGTH: 3003

TYPE: DNA

ORGANISM: Homo sapiens

US-10-044-205A-21

Query Match 74.4% Score 18.6; DB 12; Length 3003;

Best Local Similarity 84.0%; Pred. No. 7.2e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAGACAA 25
|||||
DB 1997 CCATATATATATATATCCAGACAA 2021

RESULT 10
US-10-198-846-13758

; Sequence 13758, Application US/10198846

; Publication No. US20030099974A1

GENERAL INFORMATION:

APPLICANT: Little, James

APPLICANT: Xu, Yongyao

APPLICANT: Wang, Youzhen

APPLICANT: Steinmann, Kathleen


```

1  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
2  TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
3  FILE REFERENCE: 20411-756
4  CURRENT APPLICATION NUMBER: US/09/918,995
5  CURRENT FILING DATE: 2001-07-30
6  PRIOR APPLICATION NUMBER: US/09/235,076
7  PRIOR FILING DATE: 1999-01-20
8  NUMBER OF SEQ. ID NOS: 38054
9  SOFTWARE: FastSeq for Windows Version 3.0
10 SEQ ID NO 9724
11 LENGTH: 469
12 TYPE: DNA
13 ORGANISM: Homo sapiens
14 FEATURE:
15 NAME/KEY: misc_feature
16 LOCATION: (1)...(469)
17 OTHER INFORMATION: n = A,T,C or G

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:57:09 ; Search time 12.6593 seconds
(without alignments)
3082.690 Million cell updates/sec

Title: US-09-477-082-32

Perfect score: 25

Sequence: 1 ccatatatactacatcaaaaca 25

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PC1US_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.2	76.8	7120	9	US-09-380-773-2
2	18.6	74.4	200	9	US-09-969-373-877
3	18.6	74.4	221	9	US-09-754-853A-149
4	18.6	74.4	335	10	US-09-960-352-974
5	18.6	74.4	392	10	US-09-960-352-12239
6	18.6	74.4	469	9	US-09-918-995-9724
7	18.6	74.4	487	9	US-09-918-995-13475
8	18.6	74.4	1463	10	US-09-745-763-112
9	18.6	74.4	3003	12	US-10-044-205A-21
10	18.6	74.4	4020	9	US-10-198-846-15758
11	18.6	74.4	6071	9	US-10-239-676-78
12	18.6	74.4	335913	9	US-09-754-853A-2
13	18.6	74.4	335913	9	US-09-754-853A-3
14	18.2	72.8	132	10	US-09-969-373-868
15	18.2	72.8	234	9	US-10-066-543-2507
16	18.2	72.8	259	10	US-09-969-373-861
17	18.2	72.8	259	10	US-09-969-373-862
18	18.2	72.8	746	10	US-09-770-149-33
19	18.2	72.8	822	9	US-10-144-929-55

20	18.2	72.8	1410	9	US-09-925-299-179	Sequence 179, App
21	18.2	72.8	1410	9	US-10-106-698-1413	Sequence 1413, Ap
22	18.2	72.8	1410	10	US-09-925-299-179	Sequence 179, App
23	18.2	72.8	1755	9	US-10-106-698-363	Sequence 363, App
24	18.2	72.8	3015	9	US-10-128-714-147	Sequence 147, App
25	18.2	72.8	3015	9	US-10-128-714-5147	Sequence 5147, Ap
26	18.2	72.8	5775	9	US-09-866-557A-1	Sequence 1, Appl
27	18.2	72.8	5775	9	US-10-055-797-1	Sequence 1, Appl
28	18.2	72.8	11735	9	US-10-239-676-34	Sequence 34, Appl
29	18.2	72.8	335913	9	US-09-754-853A-2	Sequence 2, Appl
30	18.2	72.8	335913	9	US-09-754-853A-3	Sequence 3, Appl
31	18.2	72.8	337628	10	US-09-813-320-3	Sequence 3, Appl
32	18.2	71.2	369	10	US-09-864-761-13372	Sequence 13372, A
33	17.6	70.4	330	9	US-09-764-891-1640	Sequence 1640, Ap
34	17.6	70.4	339	9	US-09-764-891-7638	Sequence 7638, Ap
35	17.6	70.4	339	9	US-09-764-891-7639	Sequence 7639, Ap
36	17.6	70.4	339	9	US-09-764-891-7640	Sequence 7640, Ap
37	17.6	70.4	339	9	US-09-764-891-8289	Sequence 8289, Ap
38	17.6	70.4	339	9	US-09-764-891-8290	Sequence 8290, Ap
39	17.6	70.4	349	9	US-09-764-891-8291	Sequence 8291, Ap
40	17.6	70.4	349	9	US-09-813-358-180	Sequence 180, App
41	17.6	70.4	423	10	US-09-867-701-7081	Sequence 7081, Ap
42	17.6	70.4	442	9	US-09-867-701-7081	Sequence 663, App
43	17.6	70.4	442	9	US-09-796-692-7547	Sequence 7547, Ap
44	17.6	70.4	442	9	US-09-796-692-7547	Sequence 7547, Ap
45	17.6	70.4	442	9	US-10-040-862-663	Sequence 663, App

ALIGNMENTS

RESULT 1
US-09-380-773-2/c
Sequence 2, Application US/09380773
Publication No. US20030113884A1
GENERAL INFORMATION:
APPLICANT: Heln, Silke
APPLICANT: Schilling, Brigitte
APPLICANT: Gottschalk, Gerhard
TITLE OF INVENTION: Methods for the Biosynthesis of Polyesters
FILE REFERENCE: MOBT136--- 118899, 0136, NPUS00
CURRENT APPLICATION NUMBER: US/09/380,773
CURRENT FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: PCT/US97/03994
PRIOR FILING DATE: 1997-03-02
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 7120
TYPE: DNA
ORGANISM: Clostridium kluyveri
US-09-380-773-2

Query Match 76.8%; Score 19.2; DB 9; Length 7120;
Best Local Similarity 87.5%; Pred. No. 4.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCAATATATCTCACTTCAAAACA 24
DB 3731 CCAATATATCTCACTTCAAAACA 3708

RESULT 2
US-09-969-373-877
Sequence 877, Application US/09969373
Patent No. US2002013852A1
GENERAL INFORMATION:
APPLICANT: Effertz, Roger J.
APPLICANT: Haugse, Brian M.
TITLE OF INVENTION: Soybean SSRs and Methods of Genotyping
FILE REFERENCE: 38-10(52679)A
CURRENT APPLICATION NUMBER: US/09/969, 373

```

; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brooks, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 251:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3652 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-251

```

```

Query Match          68.0%; Score 17; DB 4; Length 3652;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 CCATATATCTCATTCATAACAA 25
         ||||| ||||| ||||| |||||
Db      2085 CCATAAATATCTTGATTAAATCAA 2109

```

Search completed: July 8, 2003, 05:57:14
Job time : 22.9965 secs


```

? INFORMATION FOR SEQ ID NO: 6:
?
? SEQUENCE CHARACTERISTICS:
?     LENGTH: 2111 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: single
?     TOPOLOGY: linear
?     IMMEDIATE SOURCE:
?     LIBRARY: BRSTNOT04

```

SEQ ID NO 1
LENGTH: 11970
TYPE: DNA
ORGANISM: Homo sapiens
US-09-345-217-1

Query Match
Best Local Similarity 83.3%; Pred. No. 93;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

2 CATATATCTACATTCATAAACA 25
11995 CATACAGAGCTACATTAATAAACA 11918

RESULT 6
US-09-128-155-16
Sequence 16, Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128.155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match
Best Local Similarity 83.3%; Pred. No. 88;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

2 CATATATCTACATTCATAAACA 25
49288 CATACAGAGCTACATTAATAAACA 49311

RESULT 7
US-08-616-368A-17
Sequence 17, Application US/08616368A
Patent No. 5767262
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Haber, Edgar
APPLICANT: Jain, Mukesh
APPLICANT: Yet, Shaw-Fang
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/616,368A
FILING DATE: 15-MAR-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 05433/022001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-616-368A-17

Query Match
Best Local Similarity 86.4%; Pred. No. 1,4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

3 ATATATATCTACATTCATAAACA 24
267 ATATATATATATATACATAAACA 288

RESULT 8
US-09-054-298-17
Sequence 17, Application US/09054298
Patent No. 6136953
GENERAL INFORMATION:
APPLICANT: Lee, Mu-En
APPLICANT: Haber, Edgar
APPLICANT: Jain, Mukesh
APPLICANT: Yet, Shaw-Fang
TITLE OF INVENTION: SMOOTH MUSCLE CELL LIM PROTEIN
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/054,298
FILING DATE: 02-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/616,368
FILING DATE: 15-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Beattie, Ph.D., Ingrid A.
REGISTRATION NUMBER: P-42,306
REFERENCE/DOCKET NUMBER: 05433/022002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: both

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-Jul-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-Jan-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0980
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 1386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: INTRON 9 OF RAD50 GENOMIC SEQUENCE
US-08-687-080-76

Query Match          72.8%; Score 18.2; DB 2; Length 1386;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 CATATATATCTACATTCACAAACA 24
DB 278 CATAGAAATGTACATTCACAAACA 300

RESULT 3
US-09-853-768-3
Sequence 3, Application US/09853768
Patent No. 6444466
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF HELICASE-NOI EXPRESSION
FILE REFERENCE: RUS-0217
CURRENT APPLICATION NUMBER: US/09/853,768
CURRENT FILING DATE: 2001-05-10
NUMBER OF SEQ ID NOS: 91
SEQ ID NO 3
LENGTH: 7037
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (183)...(5957)
US-09-853-768-3

Query Match          72.8%; Score 18.2; DB 4; Length 7037;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 ATATATATCTACATTCACAAACA 25
DB 1042 ATATATCTGTACATTCACAAAGAA 1064

RESULT 4
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
```

```
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhan
APPLICANT: Williams, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: linear
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match          72.8%; Score 18.2; DB 2; Length 19124;
Best Local Similarity 87.0%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 CATATATATCTACATTCACAAACA 24
DB 7136 CATATATATATACATTCACAAACA 7157

RESULT 5
US-09-345-217-1
Sequence 1, Application US/09345217
Patent No. 6268142
GENERAL INFORMATION:
APPLICANT: DUFF, GORDON W.
APPLICANT: COX, ANGELA
APPLICANT: CAMP, NICOLA J.
APPLICANT: DIGIOVINE, FRANCESCO S.
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS FOR DISEASES ASSOCIATED
WITH AN IL-1 INFLAMMATORY HAPLOTYPE
FILE REFERENCE: MSA-010.02
CURRENT APPLICATION NUMBER: US/09/345,217
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/GB98/01481
EARLIER FILING DATE: 1998-05-21
EARLIER APPLICATION NUMBER: 9711040.7
EARLIER FILING DATE: 1997-05-29
NUMBER OF SEQ ID NOS: 32
SOFTWARE: Patentin Ver. 2.0
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:45:20 / Search time 5.99652 Seconds

(without alignments)
1278.561 Million cell updates/sec

Title: US-09-477-082-32

Sequence: 1 ccatatatactacattcaaacaa 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents, NA: *
1: /cgn2_6/prodata/1/lna/5A_COMB.seq: *
2: /cgn2_6/prodata/1/lna/5B_COMB.seq: *
3: /cgn2_6/prodata/1/lna/5A_COMB.seq: *
4: /cgn2_6/prodata/1/lna/5B_COMB.seq: *
5: /cgn2_6/prodata/1/lna/PCUTS_COMB.seq: *
6: /cgn2_6/prodata/1/lna/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	18.6	74.4	798	4	US-09-280-116-82
2	18.2	72.8	1386	2	US-08-687-080-76
3	18.2	72.8	7037	4	US-09-853-768-3
4	18.2	72.8	19124	2	US-08-487-8268-13
5	17.6	70.4	11970	4	US-09-345-217-1
6	17.6	70.4	152331	3	US-09-128-155-16
7	17.2	68.8	297	3	US-08-616-368A-17
8	17.2	68.8	297	3	US-09-054-298-17
9	17.2	68.8	297	3	US-08-818-655-17
10	17.2	68.8	342	4	US-09-134-001C-2219
11	17.2	68.8	2111	2	US-08-966-316-6
12	17.2	68.0	945	4	US-08-858-207A-111
13	17.2	68.0	1952	4	US-09-315-444-115
14	17.2	68.0	1952	4	US-09-721-362-115
15	17.2	68.0	3652	2	US-08-961-527-251
16	17.2	68.0	10754	2	US-08-966-958-1
17	17.2	68.0	10754	2	US-09-315-817-1
18	17.2	68.0	10754	2	US-09-342-353-1
19	17.2	66.4	57	4	US-08-273-594-33
20	16.6	66.4	57	3	US-09-094-919-33
21	16.6	66.4	1260	3	US-08-793-331-1
22	16.6	66.4	2763	4	US-09-489-868A-3
23	16.6	66.4	3023	4	US-09-203-453-4
24	16.6	66.4	3959	1	US-08-474-067-1
25	16.6	66.4	3959	2	US-08-474-068A-1
26	16.6	66.4	3959	2	US-08-472-481-1
27	16.6	66.4	5622	4	US-09-067-800-3

C 28	16.6	66.4	5622	4	US-09-349-677-3	Sequence 3, Appl
C 29	16.6	66.4	10007	4	US-09-410-464-13	Sequence 13, Appl
C 30	16.6	66.4	70000	4	US-09-851-896-3	Sequence 3, Appl
C 31	16.6	66.4	72604	4	US-09-268-992-7	Sequence 7, Appl
C 32	16.6	66.4	72604	4	US-09-657-474-7	Sequence 7, Appl
C 33	16.6	66.4	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 34	16.6	66.4	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 35	16.6	66.4	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 36	16.4	65.6	435	4	US-09-134-001C-1206	Sequence 1206, Ap
C 37	16.4	65.6	10614	1	US-08-135-511-35	Sequence 35, Appl
C 38	16.4	65.6	10614	1	US-08-187-453-35	Sequence 35, Appl
C 39	16.2	64.8	470	4	US-09-221-017B-324	Sequence 324, App
C 40	16.2	64.8	685	4	US-08-822-999-10	Sequence 10, Appl
C 41	16.2	64.8	693	2	US-08-791-347-16	Sequence 16, Appl
C 42	16.2	64.8	2000	4	US-09-056-285A-3	Sequence 3, Appl
C 43	16.2	64.8	2004	4	US-08-961-527-95	Sequence 95, Appl
C 44	16.2	64.8	2606	4	US-09-252-816A-7	Sequence 7, Appl
C 45	16.2	64.8	2981	1	US-08-257-073-2	Sequence 2, Appl

ALIGNMENTS

```

RESULT 1
US-09-280-116-82
; Sequence 82, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 82
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: matrix metalloproteases
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(798)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-82

Query Match          74.4%; Score 18.6; DB 4; Length 798;
Best Local Similarity 84.0%; Pred. No. 40;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCAATATATCTCATTCACAAACAA 25
    |||||  |||||  |||||  |||||
DB 422 CCAATATATATATATCAAGACAA 446

RESULT 2
US-08-687-080-76
; Sequence 76, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```


REFERENCE 3 (bases 1 to 923)
 AUTHORS Genoscope
 JOURNAL Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvaurum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
lactis var. *lactis*, *Kluyveromyces fragilis* var. *fragilis*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 source
 1..923
 /organism="Pichia farinosa"
 /strain="CBS 7064"
 /db_xref="taxon:4920"
 /clone="AX0AA030D10"
 /clone_1lb="AX0AA"
 /note="end : T3"
 <10..>150
 /note="similar to *Saccharomyces cerevisiae* ORF YBR217W [AP012 ; component of the autophagic system]"
 /evidence=not-experimental
 complement(<652..>918)
 /note="similar to *Saccharomyces cerevisiae* ORF YJL123W [SIM1 ; involved in cell cycle regulation and aging]
 similar to *Saccharomyces cerevisiae* ORF YKR042W [UTH1 ; involved in the aging process]
 similar to *Saccharomyces cerevisiae* ORF YNL066W [SUN4 ; cell wall protein with homology to beta-glucosidase]"
 /evidence=not-experimental
 complement(<670..>915)
 /note="similar to *Saccharomyces cerevisiae* ORF YJL116C [NCA3 ; involved in regulation of synthesis of Atp8p and Atp8p]"
 /evidence=not-experimental

misc_feature
 326 a 203 c 156 g 236 t 2 others
 BASE COUNT 326 a 203 c 156 g 236 t 2 others
 ORIGIN

Query Match 77.6% ; Score 19.4 ; DB 17 ; Length 923 ;
 Best Local Similarity 95.2% ; Pred. No. 1.4e+03 ;
 Matches 20 ; Conservative 0 ; Mismatches 1 ; Indels 0 ; Gaps 0 ;

QY 3 ATATATATCTACATTCACAAC 23
 ||||| ||||| ||||| |||||
 Db 28 ATATATGCTCTACATTCACAAC 48

RESULT 15
 CDS06GCM
 LOCUS T3 end of clone AX0AA030D10 of library AX0AA from strain CBS 7064
 DEFINITION of *Pichia farinosa*, genomic survey sequence.
 ACCESSION AL418278
 VERSION AL418278.1 GI:12200835
 KEYWORDS GSS.
 SOURCE *Pichia farinosa*.
 ORGANISM *Pichia farinosa*.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; *Pichia*.
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
 de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
 Malpertuy,A., Neugebaur,C., Ozler-Kalogeropoulos,O., Potier,S.,
 Saurin,W., Tekala,F., Toffano-Nicche,C., Wesolowski-Louvel,M.,
 Winkler,P. and Weissbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of

JOURNAL Yeast species for molecular evolution studies
 MEDLINE FEMS Lett. 487 (1), 3-12 (2000)
 PUBMED 20584711
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 965)
 AUTHORS de Montigny,J., Spohner,C., Souciet,J., Tekala,F., Dujon,B.,
 Winkler,P., Artiguenave,F. and Potier,S.
 COMMENT Genomic exploration of the hemiascomycetous yeasts: 15. *Pichia*
sorbitophila
 FEMS Lett. 487 (1), 87-90 (2000)
 JOURNAL MEDLINE 20584725
 PUBMED 11152890
 REFERENCE 3 (bases 1 to 965)
 AUTHORS Genoscope.
 JOURNAL Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvaurum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
lactis var. *lactis*, *Kluyveromyces fragilis* var. *fragilis*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES
 source
 1..965
 /organism="Pichia farinosa"
 /strain="CBS 7064"
 /db_xref="taxon:4920"
 /clone="AX0AA030D10"
 /clone_1lb="AX0AA"
 /note="end : T3"
 <10..>150
 /note="similar to *Saccharomyces cerevisiae* ORF YBR217W [AP012 ; component of the autophagic system]"
 /evidence=not-experimental
 complement(<652..>942)
 /note="similar to *Saccharomyces cerevisiae* ORF YJL123W [SIM1 ; involved in cell cycle regulation and aging]
 similar to *Saccharomyces cerevisiae* ORF YKR042W [UTH1 ; involved in the aging process]
 similar to *Saccharomyces cerevisiae* ORF YNL066W [SUN4 ; cell wall protein with homology to beta-glucosidase]"
 /evidence=not-experimental
 complement(<670..>942)
 /note="similar to *Saccharomyces cerevisiae* ORF YJL116C [NCA3 ; involved in regulation of synthesis of Atp8p and Atp8p]"
 /evidence=not-experimental

misc_feature
 339 a 214 c 164 g 247 t 1 others
 BASE COUNT 339 a 214 c 164 g 247 t 1 others
 ORIGIN

Query Match 77.6% ; Score 19.4 ; DB 17 ; Length 965 ;
 Best Local Similarity 95.2% ; Pred. No. 1.4e+03 ;
 Matches 20 ; Conservative 0 ; Mismatches 1 ; Indels 0 ; Gaps 0 ;

QY 3 ATATATATCTACATTCACAAC 23
 ||||| ||||| ||||| |||||
 Db 28 ATATATGCTCTACATTCACAAC 48

Search completed: July 8, 2003, 05:51:43
 Job time : 193.499 secs

REFERENCE 1 (bases 1 to 651)
AUTHORS Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P., and Venter, J.C.
TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
COMMENT Unpublished (1997)
Other_GSSs: RPCI11-12117.TV
Contact: Shaying Zhao, William Niernan, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hdelety@ig.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics (<http://info@resgen.com>). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1..651
/organism="Homo sapiens"
/db_xref="GDB:7546278"
/db_xref="taxon:9606"
/clone="RPCI-11-12117"
/clone_11b="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
BASE COUNT 213 a 81 c 96 g 261 t
ORIGIN

Query Match 77.6%; Score 19.4; DB 17; Length 651;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATATATCTACATTCATAAC 23
|||||
Db 296 ATATATGTACATTCATAAC 276

RESULT 13
CNS060RT 849 bp DNA linear GSS 06-JUL-2001
LOCUS T3 end of clone AX0AA012F02 of library AX0AA from strain CBS 7064
ACCESSION AL416212
VERSION AL416212.1 GI:12196387
KEYWORDS GSS.
SOURCE Pichia farinosa.
ORGANISM Pichia farinosa.
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
1 (bases 1 to 849)
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Ben, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Maupertuy, A., Nevegilise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 849)
AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F., and Potier, S.

TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila
JOURNAL FEBS Lett. 487 (1), 87-90 (2000)
MEDLINE 20584725
PUBMED 11152890
REFERENCE 3 (bases 1 to 849)
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Ben, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Maupertuy, A., Nevegilise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 923)
AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F., and Potier, S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila
JOURNAL FEBS Lett. 487 (1), 87-90 (2000)
MEDLINE 20584725
PUBMED 11152890

FEATURES
source 1..849
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0AA012F02"
/clone_11b="AX0AA"
/note="end : T3"
Location/Qualifiers
1..849
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0AA012F02"
/clone_11b="AX0AA"
/note="end : T3"
BASE COUNT 280 a 167 c 150 g 242 t 10 others
ORIGIN

Query Match 77.6%; Score 19.4; DB 17; Length 849;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATATATCTACATTCATAAC 23
|||||
Db 553 ATATATGTACATTCATAAC 573

RESULT 14
CNS060MET 923 bp DNA linear GSS 06-JUL-2001
LOCUS T3 end of clone AX0AA030H08 of library AX0AA from strain CBS 7064
ACCESSION AL418347
VERSION AL418347.1 GI:12200957
KEYWORDS GSS.
SOURCE Pichia farinosa.
ORGANISM Pichia farinosa.
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
1 (bases 1 to 923)
AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Ben, E., Brothier, P., Casaregola, S., de Montigny, J., Dujon, B., Durans, P., Lepingle, A., Llorente, B., Maupertuy, A., Nevegilise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 923)
AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F., and Potier, S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila
JOURNAL FEBS Lett. 487 (1), 87-90 (2000)
MEDLINE 20584725
PUBMED 11152890

FEATURES
source
Location/Qualifiers
1. 764
/organism="Brassica oleracea"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMH023"
/clone_lib="BO_2-3_KB"
/note="Vector: pHC01; Site_1: BstXI; 2-3 kb sheared genomic DNA inserted into pHC01 using BstXI linkers"

BASE COUNT
223 a 140 c 171 g 230 t

ORIGIN
Query Match 79.2%; Score 19.8; DB 17; Length 764;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 ATATATCTACATTCAAAACA 25
|||||
Db 723 ATATATCTACATTCAAAATA 745

RESULT 10
CNS04K6T 793 bp DNA linear GSS 21-MAY-2000
LOCUS
DEFINITION
Tetradodon nigroviridis genome survey sequence T7 end of clone 115P02 of library G from Tetradodon nigroviridis, genomic survey sequence.
ACCESSION
AL294446.1 GI:8033026
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetradodon nigroviridis.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradodon.
1 (bases 1 to 793)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetradodon nigroviridis DNA sequence

TITLE
Unpublished
2 (bases 1 to 793)
Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis

JOURNAL
Unpublished
3 (bases 1 to 793)
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (12-APR-2000)
COMMENT
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradodon>.
Location/Qualifiers
1. 793
/organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone="115P02"
/clone_lib="G"
/note="Genoscope sequence ID: COBGI15DH01.P1-end : T7"

BASE COUNT
220 a 94 c 208 g 240 t 31 others

ORIGIN
Query Match 79.2%; Score 19.8; DB 17; Length 793;
Best Local Similarity 84.0%; Pred. No. 1e+03; 3; Indels 0; Gaps 0;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 CCATATATCTACATTCAAAACA 25

Db 198 CCATATATCTCTCATTCAAAACA 174
|||||
RESULT 11
A2519839/C 513 bp DNA linear GSS 16-OCT-2000
LOCUS
DEFINITION
RPCI-11-79F2.TVB RPCI-11 Homo sapiens genomic clone RPCI-11-79F2, DNA sequence.
ACCESSION
A2519839.1 GI:10831743
VERSION
A2519839.1
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 513)
Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
J.C.
BAC end sequences of library RPCI-11
Unpublished (1997)
Other GSSs: RPCI11-79F2.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Plietier de Jong (plietier@jg.med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from Research Genet cs (<http://infogenet.com>). BAC end search page: http://www.tigr.org/tdb/human/bac/end_search/bac_end_search.html. This BAC end was generated during the Rdb process and may have higher chance of clone tracking errors.
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
1. 513
/organism="Homo sapiens"
/db_xref="GDB:7530873"
/db_xref="taxon:9606"
/clone="RPCI-11-79F2"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT
168 a 65 c 76 g 203 t 1 others

ORIGIN
Query Match 77.6%; Score 19.4; DB 17; Length 513;
Best Local Similarity 95.2%; Pred. No. 1.4e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 ATATATCTACATTCAAAAC 23
|||||
Db 269 ATATATCTACATTCAAAAC 249

RESULT 12
AQ342956/C 651 bp DNA linear GSS 07-MAY-1999
LOCUS
DEFINITION
RPCI11-12117.TV RPCI-11 Homo sapiens genomic clone RPCI-11-12117, DNA sequence.
ACCESSION
AQ342956.1 GI:4167852
VERSION
AQ342956.1
KEYWORDS
GSS.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawachi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES
source
Location/Qualifiers
1..369
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Clontech human aorta polyA+ mRNA (#5572)"
/note="Male adult, hematopoietic tissue, stem cell"

BASE COUNT
ORIGIN
108 a 54 c 57 g 140 t 10 others

Query Match 79.2%; Score 19.8; DB 14; Length 369;
Best Local Similarity 84.0%; Pred. No. 1e+03;
Matches 21; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATCTACATCAACA 25
DB 79 CCATATATATATCAACA 103

RESULT 7
AA966215/c 371 bp mRNA linear EST 31-JUL-1998
LOCUS
DEFINITION
v4e09a1.f1 Aspergillus nidulans 24hr asexual developmental and
vegetative cDNA lambda zap library Emericella nidulans cDNA clone
v4e09a1 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AA966215
AA966215.1 GI:3140297
EST.
Emericella nidulans.
Emericella nidulans.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiiales; Trichocomaceae; Emericella.
1 (bases 1 to 371)
Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,
Prade,R. and Roe,B.
An Aspergillus nidulans EST Database
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: M13-20
High quality sequence stop: 208.

FEATURES
source
Location/Qualifiers
1..371
/organism="Emericella nidulans"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clone_lib="v4e09a1"
/clone_lib="Aspergillus nidulans 24hr asexual
developmental and vegetative cDNA lambda zap library"
/issue_type="vegetative mycelia, asexual structures"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT
ORIGIN
129 a 60 c 106 g 76 t

Query Match 79.2%; Score 19.8; DB 9; Length 371;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 CATATATCTACATCAACA 24
||||||| ||||| |||||

DB 47 CATATATATCTACATCAACA 25

RESULT 8
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bj401889
Bj401889 Dictyostelium discoideum cDNA library, SF Dictyostelium
discoideum cDNA clone dds19a20 3', mRNA sequence.
Bj401889
Bj401889.1 GI:19314806
EST.
Dictyostelium discoideum.
Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
1 (bases 1 to 614)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
Full length cDNA of Dictyostelium discoideum at the slug stage
Unpublished (2002)
Contact: Tadasi Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsuhit@genes.nig.ac.jp.

FEATURES
source
Location/Qualifiers
1..614
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dds19a20"
/clone_lib="Dictyostelium discoideum cDNA library, SF"
/sex="mat A"
/dev_stage="Slug stage"

BASE COUNT
ORIGIN
244 a 65 c 70 g 234 t 1 others

Query Match 79.2%; Score 19.8; DB 13; Length 614;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATATATCTACATCAACA 25
DB 503 ATATATATATCAACA 481

RESULT 9
BH701475 764 bp DNA linear GSS 20-FEB-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BOMH023TR BO_2_3_KB Brassica oleracea genomic clone BOMH023, DNA
sequence.
BH701475
BH701475.1 GI:18779250
GSS.
Brassica oleracea.
Brassica oleracea.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 764)
Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other-GSSs: BOMH023TR
Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagacka, S., Sasakawa, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermolabile and thermolabile enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kusunagi, T., Akiyama, Y., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES

Source

1. 318
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C330036108"
/clone_1ib="RIKEN full-length enriched, ES cells"
/cell_type="ES cells"
/lab_host="SOLR"
/note="Site_1: XhoI; site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGAGACCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 5.0, and subtraction to Rot = 25.0. Second strand
cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGATTCGAGATTAAATTAATTAATCCCCCCCCCC 3']"

BASE COUNT

96 a 55 c 39 g 128 t

ORIGIN

Query Match 79.2%; Score 19.8; DB 10; Length 318;
Best Local Similarity 91.3%; Pred. No. 1e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

QY 3 ATATATCTACATCAAAACAA 25
|||||
Db 220 ATATTATCTACATTTAAACAA 242

RESULT 6

D62159 369 bp mRNA linear EST 29-AUG-1995
LOCUS HUM242G12B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens
DEFINITION cDNA clone GEN-242G12 5', mRNA sequence.
ACCESSION D62159
VERSION D62159
KEYWORDS GI:965935
SOURCE EST.
ORGANISM human.
Homosapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Fujisawa, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shimomiyu, H., Takaiuchi,
A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H.,
Shin, S. and Nakamura, Y.
Fujisawa et al. (1995)
TITLE unpublished
JOURNAL Contact: Tsutomu Fujisawa
COMMENT Otsuka GEN Research Institute

Email: acameron@caltech.edu
 Plate: 48 row: L column: 2

Seq primer: 17
 Class: BAC ends
 High quality sequence stop: 504.
 Location/Qualifiers

FEATURES

source

1..504
 /organism="Strongylocentrotus purpuratus"
 /db_xref="taxon:7668"
 /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
 /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Clon DH10B"
 BASE COUNT 153 a 100 c 74 g 171 t 6 others
 ORIGIN

Query Match 81.6%; Score 20.4; DB 17; Length 504;
 Best Local Similarity 95.5%; Pred. No. 5.9e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATATATCTACATTCAAACA 24
 |||||
 116 ATATATATACATTCAAACA 95

RESULT 2
 BE664368/c 127 bp mRNA linear EST 25-APR-2001
 LOCUS 148797 MARC 4BOV Bos taurus cDNA 5', mRNA sequence.
 DEFINITION BE664368
 ACCESSION BE664368.1 GI:10023781
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

REFERENCE
 1 (bases 1 to 127)
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitto-McKown,C.G.,
 Percec,G., Holt,I., Karanicheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.

TITLE
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 JOURNAL
 MEDLINE
 2118013
 Genome Res. 11 (4), 626-630 (2001)

COMMENT
 Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390

Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980304.e. Vector identified by cross_match with the -minscore 18
 and mismatch 12 options.

PCR Primers
 FORWARD: AGGAACACGATGACACAT
 BACKWARD: GTTCCAGTCACGACG
 Plate: 52 row: L column: 10
 Seq primer: ATTAGTGACACTARAG.

FEATURES

source

1..127
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 4BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site_1: NotI; Site_2: SalI;
 Library made from pooled tissue from day 20 and day 40
 embryos."

BASE COUNT

40 a 18 c 26 g 43 t

ORIGIN

Query Match 80.8%; Score 20.2; DB 10; Length 127;
 Best Local Similarity 88.0%; Pred. No. 7.1e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATCTACATTCAAACA 25
 |||||
 52 CCAATATATCTACATTCAAACA 28

RESULT 3
 BH082823/c 398 bp DNA linear GSS 18-JUL-2001
 LOCUS RPCI-24-236A11.TJ RPCI-24 Mus musculus genomic clone RPCI-24-236A11
 DEFINITION
 , DNA sequence.
 ACCESION BH082823
 VERSION BH082823.1 GI:14902420
 KEYWORDS
 SOURCE
 ORGANISM Mus musculus
 house mouse.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 398)
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-236A11.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@tigr.org

REFERENCE
 1 (bases 1 to 398)
 Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akiret,B., Levins,M.,
 Tsegaye,G., Geer,K., Krol,M., Shwartsbeyn,A., Gebregeorgis,E.,
 Russell,D., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-24
 Unpublished (1999)
 Other_GSSs: RPCI-24-236A11.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhaoc@tigr.org

TITLE
 JOURNAL
 COMMENT
 Clones are derived from the mouse BAC library RPCI-24. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@tigr.org). Clones may be purchased from BACPAC
 Resources (<http://www.chori.org/bacpac/orderframe.htm>). BAC end
 page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 236 row: A column: 11
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source

1..398
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone_RPCI-24-236A11"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBA1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
 library was cloned in the pTARBA1 cloning vector at the
 BamHI sites using MboI partially digested male C57BL/6J
 DNA."

BASE COUNT 144 a 47 c 84 g 123 t

Query Match 80.8%; Score 20.2; DB 17; Length 398;
 Best Local Similarity 88.0%; Pred. No. 7.1e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATCTACATTCAAACA 25
 |||||
 249 CCTATATATCTACATTCAAATAA 225

RESULT 4
 DRIP18T/c

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 02:04:24 ; Search time 188.499 seconds

(without alignments)
2147.949 Million cell updates/sec

Title: US-09-477-082-32

Perfect score: 25
Sequence: 1 cccatatactcattcaaacaa 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estdb: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_estl: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	20.4	81.6	504	17	AZ149876
2	20.2	80.8	127	10	BE664368
3	20.2	80.8	398	17	BH082823
4	20.2	80.8	582	17	DR1P18T
5	19.8	79.2	318	10	BB404480
6	19.8	79.2	369	14	D62159

C	7	19.8	79.2	371	9	AA966215	AA966215 v4e09a1.f
C	8	19.8	79.2	614	13	BH401889	BH401889 BH401889
C	9	19.8	79.2	764	17	BH701475	BH701475 BOMH023TF
C	10	19.8	79.2	793	17	CNS04K6T	AL294446 Tetradon
C	11	19.4	77.6	513	17	A2519839	A2519839 RPT-11-7
C	12	19.4	77.6	651	17	AQ342956	AQ342956 RPT-11-12
C	13	19.4	77.6	849	17	CNS06UR1	AL416212 T3 end of
C	14	19.4	77.6	923	17	CNS06WET	AL418447 T3 end of
C	15	19.4	77.6	965	17	CNS06MCW	AL418478 T3 end of
C	16	19.2	76.8	265	13	BJ052358	BJ052358 BJ052358
C	17	19.2	76.8	312	9	AA661350	AA661350 MBACFW3A0
C	18	19.2	76.8	328	17	TA58F050	AL455735 T. brucei
C	19	19.2	76.8	356	13	BM026941	BM026941 GIT000020
C	20	19.2	76.8	439	9	AA509210	AA509210 MBACFX8F0
C	21	19.2	76.8	466	14	N74857	N74857 SW3ICA1435S
C	22	19.2	76.8	506	9	AA585509	AA585509 SW3D9CA05
C	23	19.2	76.8	511	13	BM027113	BM027113 GIT000040
C	24	19.2	76.8	549	17	A2317047	A2317047 1M0035H11
C	25	19.2	76.8	581	17	AQ655892	AQ655892 Sheared D
C	26	19.2	76.8	582	10	AM675599	AM675599 SWID25CAU
C	27	19.2	76.8	586	9	A1784855	A1784855 SWACAC30
C	28	19.2	76.8	654	10	AM874758	AM874758 SWYACAL06
C	29	19.2	76.8	655	17	BH768439	BH768439 BMBAC360A
C	30	19.2	76.8	691	17	AG177614	AG177614 Pac trogl
C	31	19.2	76.8	702	17	BH476823	BH476823 BOCIV58TF
C	32	19.2	76.8	704	17	BH033807	BH033807 RPT-24-2
C	33	19.2	76.8	762	17	BH420628	BH420628 BOMOR5TR
C	34	19.2	76.8	792	13	BI196245	BI196245 602754702
C	35	19.2	76.8	864	17	AZ203406	AZ203406 SP. 0079_B
C	36	19.2	76.8	1494	17	AG043850	AG043850 Pan trogl
C	37	19.2	76.0	1061	17	CNS020MA	AL209467 Tetradon
C	38	18.8	75.2	304	9	A1110986	A1110986 MF002D05
C	39	18.8	75.2	341	10	BE648998	BE648998 UT-M-BH2
C	40	18.8	75.2	395	9	AU227314	AU227314 AU227314
C	41	18.8	75.2	405	10	AM414891	AM414891 48889 MAR
C	42	18.8	75.2	418	12	BF850427	BF850427 CM0-ENO10
C	43	18.8	75.2	433	17	AQ215510	AQ215510 HS.2259_B
C	44	18.8	75.2	445	9	A1110889	A1110889 MF003D10
C	45	18.8	75.2	445	12	BE953439	BE953439 UT-M-CD1-

ALIGNMENTS

RESULT 1
AZ149876/c 504 bp DNA linear GSS 28-AUG-2000
SP_0048_B2_F01_T7 Strongylocentrotus purpuratus, purple sea urchin,
sperm genomic BAC library Strongylocentrotus purpuratus genomic
clone plate=48 Col=2 Row=L, DNA sequence.

ACCESSION AZ149876
VERSION AZ149876.1 GI:8301777
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus.

REFERENCE
AUTHORS Cameron,R.A., Mahaliras,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Poultka,A.J., Livingston,B.T., Wray
,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and
Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)

TITLE JOURNAL
MEDLINE
COMMENT
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047

PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259676.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-483426/52.
 DR P-PSDB; AAM88893.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Claim 1; SEQ ID NO 6734; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SO Sequence 379 BP; 108 A; 62 C; 104 G; 101 T; 4 other;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 CCATATATATCTACATTCAAACAA 25
 ||| ||||| ||||| |||||
 Db 209 CCACATATATCTTCATTAAAAACCA 185

Search completed: July 8, 2003, 03:11:19
 Job time : 27.9421 secs

Query Match 74.4%; Score 18.6; DB 22; Length 379;
 Best Local Similarity 84.0%; Pred. No. 3.3e+02;

CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA03910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapies. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoietic regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 361 BP; 100 A; 52 C; 73 G; 136 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 361;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 CCATATATCTCATCTCAACAA 25
D6 175 CTATATATATTTCATTCACAAA 151

RESULT 15

AAK61674/c
ID AAK61674 standard; CDNA; 379 BP.

XX AAK61674;

DT 06-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen encoding CDNA SEQ ID NO:6734.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metastasis; ss.

XX Homo sapiens.

PN WO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0198874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214866.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218299.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0234584.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIDENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

PS Claim 1; SEQ ID NO 35; 32bp + Sequence Listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

CC Sequence 9905 BP; 3133 A; 75 C; 1920 G; 4777 T; 0 other;

Query Match 75.2%; Score 18.8; DB 24; Length 9905;

Best Local Similarity 90.9%; Pred. No. 3e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATATATATCTACTTCAAAACA 24
 ||||||| ||||||| ||

DB 7931 ATATATATATACATTCAAACCA 7910

RESULT 13

AA161518 ID AA161518 standard; DNA; 221 BP.

AC AA161518;

DT 16-OCT-2001 (first entry)

DE Soybean 240017 region G3 DNA, SEQ ID NO: 149.

KW Soybean; antihelminthic; gene therapy; soybean cyst nematode; SCN;

KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;

KW 240017 region G3; 318013 region A3; 515002 region G2; ds.

OS Glycine max.

PN WO200151627-A2.

PD 19-JUL-2001.

PF 05-JAN-2001; 2001WO-US00552.

PR 07-JAN-2000; 2000US-0174880.

PA (MONS) MONSANTO CO.
 PI Hauge BM, Wang ML, Parsons JD, Farnell LD;
 DR WPI: 2001-425872/45.

PT New purified nucleic acid for producing a soybean plant having soybean

PT cyst nematode resistance and for use in plant breeding programs -

PS Claim 25; Page 1062; 1353pp; English.

CC The invention relates to nucleic acid molecules from regions of the
 CC soybean genome which are associated with soybean cyst nematode (SCN)
 CC resistance. The nucleic acids are used to transform plants, and can
 CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
 CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
 CC of soybean plants and for introgressing SCN resistance or partial SCN
 CC resistance into soybean plants. They can also be used in plant breeding
 CC programmes. The invention also relates to proteins encoded by such
 CC nucleic acid molecules, as well as antibodies capable of recognising
 CC these proteins. The present sequence is a nucleic acid molecule
 CC provided in the specification.

CC Sequence 221 BP; 72 A; 28 C; 28 G; 93 T; 0 other;

Query Match 74.4%; Score 18.6; DB 22; Length 221;

Best Local Similarity 84.0%; Pred. No. 3.3e+02;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCATATATATCTACTTCAAAACA 25
 ||||||| ||||| |||||||

DB 120 CGATATATATATACATTAATAAACA 144

RESULT 14

AA191681/C ID AA191681 standard; CDNA; 361 BP.

AC AA191681;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 11741.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 nervous system disorders; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577405.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR P-PSDB; AAC11750.

PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

PS Claim 1; SEQ ID NO 11741; 1399pp + Sequence Listing; English.

Query Match 76.8%; Score 19.2; DB 24; Length 15373;
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CATATATATCTACATTCATAACAA 25
 DB 12222 CTTAAATATCTACATTCATAACAA 12199

RESULT 10
 ABL33519/C
 ID ABL33519 standard; DNA: 5520 BP.

AC ABL33519;
 XX
 XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 1492.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

PS Claim 1; SEQ ID NO 1492; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 5520 BP; 1725 A; 36 C; 1004 G; 2753 T; 2 other;

Query Match 75.2%; Score 18.8; DB 24; Length 5520;
 Best Local Similarity 90.9%; Pred. No. 2.9e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TATATATCTACATTCATAACAA 25
 DB 2604 TAAATATCTACATTCATAACAA 2583

RESULT 11
 ABL32705/C

ID ABL32705 standard; DNA: 6297 BP.

XX ABL32705;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 678.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

PS Claim 1; SEQ ID NO 678; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

SQ Sequence 6297 BP; 1577 A; 217 C; 1471 G; 3032 T; 0 other;

Query Match 75.2%; Score 18.8; DB 24; Length 6297;
 Best Local Similarity 90.9%; Pred. No. 3e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TATATATCTACATTCATAACAA 25
 DB 480 TAAATATCTACATTCATAACAA 459

RESULT 12

ABL32062/C

ID ABL32062 standard; DNA: 9905 BP.

XX ABL32062;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 35.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 4654 BP; 1313 A; 50 C; 955 G; 2336 T; 0 other;
 Query Match 76.8%; Score 19.2; DB 24; Length 4654;
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CATATATATCTACATCAAAACAA 25
 ||||| ||||| ||||| ||||| |||||
 Db 874 CATATTATCTACATCAAAACAA 851

RESULT 8
 AAS46544/c
 ID AAS46544 standard; DNA; 8845 BP.
 AC AAS46544;
 XX
 AC AAS46544;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #266.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PE 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer -
 XX
 PS Claim 1; SEQ ID No 266; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and

CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 8845 BP; 2565 A; 135 C; 1694 G; 4451 T; 0 other;
 Query Match 76.8%; Score 19.2; DB 22; Length 8845;
 Best Local Similarity 87.5%; Pred. No. 2.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 CATATATATCTACATCAAAACAA 25
 ||||| ||||| ||||| ||||| |||||
 Db 8144 CAATATATCAACATCAAAACAA 8121

RESULT 9
 ABL32466/c
 ID ABL32466 standard; DNA; 15373 BP.
 AC ABL32466;
 XX
 AC ABL32466;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 439.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cyrostatic; nootropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmologic;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PE 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 439; 32pp + Sequence listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 15373 BP; 5235 A; 89 C; 2576 G; 7473 T; 0 other;

CC Insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (AB357737-AB172072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 4025 BP; 1049 A; 1020 C; 933 G; 1023 T; 0 other;

Query Match 76.8%; Score 19.2; DB 23; Length 4025;
Best Local Similarity 87.5%; Pred. NO. 2.1e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0

OY 2 CATATATATCTACATTCAAACCA 25
Db 3876 CATATATATTTTCAATGCAAAACCA 3853
|||||

RESULT 6
AAS46781/c
ID AAS46781 standard; DNA; 4654 BP.
XX
AC AAS46781;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #506.
XX
KW Human: tumour suppressor gene; oncogene; antitumour; cytostatic;
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
XX
OS Homo sapiens.
XX
PN M0200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP02955.
XX
PR 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPiG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer
XX
PS Claim 1; SEQ ID No 506; 27pp; English.

The invention relates to a nucleic acid comprising a sequence of 18
bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
bisulphite, of genes associated with tumour suppression and
oncogenes having a sequence taken from 536 (actually 533 since
numbers 408, 458 and 500 are missing from the sequence listing) sequences
(Ss) and sequences complementary to (Ss). The nucleic acid may be a
peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
form part of a set of probes for detecting the cytosine methylation state
of a single nucleotide polymorphisms and also to be used in an
array for analysing diseases associated with CpG dinucleotides e.g.
cancers and tumours. The probes can also be used in a method for
ascertaining genetic and/or epigenetic parameters for the diagnosis
and/or therapy of existing diseases or the predisposition to specific

CC	diseases by analysing cytosine methylations. The parameters may be
CC	compared to another set of genetic and/or epigenetic parameters, the
CC	differences serving as basis for diagnosis and/or prognosis events which
CC	are disadvantageous to patients. The present sequence is one of the
CC	533 genomic sequences derived from tumour suppressor genes and
CC	oncogenes. Sequences with even numbered Seq ID numbers are the
CC	complementary sequence of the corresponding odd numbered sequence (e.g.
CC	ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
CC	is missing)
CC	Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic
CC	format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences.
SO	Sequence 4654 BP; 1313 A; 50 C; 955 G; 2336 T; 0 other;
Qy	Query Match 76.8%; Score 19.2; DB 22; Length 4654;
Qy	Best Local Similarity 87.5%; Pred. No. 2,1e+02;
Qy	Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0
Db	2 CATATATATCTACATCAAAACAA 25
Db	874 CATATTTATCTACATCCAAACCA 851
RESULT 7	
ID	ABL34223/c
AC	ABL34223 standard; DNA; 4654 BP.
XX	
AC	ABL34223:
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Human immune system associated gene SEQ ID NO: 2196.
XX	
KN	Human: immune system disease; cytosine methylation; antiasthmatic;
KN	antiarteriosclerotic; antihaemic; cytosolic; noctropic;
KN	neuroprotective; anti-HIV; anticoagulant; ophthalmological;
KN	antiinfective; antiarthritic; antidiabetic; antiparasitic;
KN	antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KN	acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
KN	neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KN	gene; ds.
OS	
XX	Homo sapiens.
XX	
PN	WO200200928-A2.
XX	
PD	03-JAN-2002.
XX	
PF	02-JUL-2001; 2001WO-EP07537.
XX	
PR	30-JUN-2000; 2000DE-1032529.
PR	01-SEP-2000; 2000DE-1043826.
XX	
PA	(EP1G-) EPIGENOMICS AG.
XX	
PI	Olek A, Piepenbrock C, Berlin K;
DR	WPI: 2002-130909/17.
XX	
PT	Nucleic acid comprising fragment of chemically modified gene, useful
PT	for diagnosis and treatment of diseases associated with abnormal,
PT	cytosine methylation
XX	
PS	Claim 1: SEQ ID NO 2196; 32bp + Sequence Listing; German.
XX	
CC	The present invention provides a number of human immune system associated
CC	genes which are modified by the methylation of cytosines. The sequences
CC	can be used in the diagnosis and treatment of immune system disorders,
CC	including eye diseases such as retinopathy, neovascular glaucoma and
CC	macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC	leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

PR 18-JUL-2000; 2000US-219007P.
 PR 13-DEC-2000; 2000US-255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.
 XX
 PT Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer; useful
 PT for detecting presence of prostate cancer; stage of prostate cancer -
 XX
 PS Claim 1; Page 9866; 11750pp; English.
 XX
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement; (I) is useful for:
 CC (a) assessing whether a patient is afflicted with prostate cancer;
 CC (b) monitoring the progression of prostate cancer in a patient;
 CC (c) assessing the efficacy of a test compound to inhibit prostate
 CC cancer in a patient;
 CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
 CC in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound;
 CC (g) determining whether prostate cancer has metastasized in a patient;
 CC (h) assessing the aggressiveness or indolence of prostate cancer in a
 CC patient;
 CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
 XX
 SQ Sequence 523 BP; 173 A; 108 C; 90 G; 151 T; 1 other:
 XX
 QY Query Match 79.2%; Score 19.8; DB 23; Length 523;
 Best Local Similarity 91.3%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 DB 3 ATATATCTACATTCATAAACA 25
 411 AATATATCTACATTCATAAACA 389
 XX
 RESULT 4
 ABO67066/C
 ID ABO67066 standard; DNA; 5474 BP.
 XX
 AC ABO67066;
 XX
 DT 28-AUG-2002 (first entry)
 XX
 DE Human angiogenesis associated polynucleotide SEQ ID NO 96.
 XX
 KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiarteriosclerotic; ds.
 XX
 PA Homo sapiens.
 XX
 OS Homo sapiens.
 XX
 PI WO200246454-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 06-DEC-2001; 2001WO-EP14320.
 XX
 PR 06-DEC-2000; 2000DE-1061338.
 XX
 PS (EPIC-) EPIGENOMICS AG.
 XX
 PI Schacht O;
 XX
 DR WPI; 2002-500450/53.

XX
 PT New nucleic acid fragments from chemically treated
 PT angiogenesis-associated genes, useful for determining methylation
 PT status, e.g. in diagnosis or treatment of cancer -
 XX
 PS Claim 1; SEQ ID NO 96; 41pp + Sequence Listing; German.
 XX
 CC The invention relates to a nucleic acid (I) comprising a segment of 18
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
 CC having sequences (ABO66971-ABO67178) or their complements. (I), also
 CC related oligomers, are used to evaluate the methylation status and/or
 CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
 CC diagnosis and treatment of eye diseases, proliferative retinopathy,
 CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
 CC diabetic retinopathy, macular degeneration caused by neovascularisation,
 CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
 CC Crohn's disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 5474 BP; 1440 A; 107 C; 1627 G; 2300 T; 0 other:
 XX
 QY Query Match 79.2%; Score 19.8; DB 24; Length 5474;
 Best Local Similarity 91.3%; Pred. No. 1.2e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 XX
 DB 2 CATATATCTACATTCATAAACA 24
 1823 CATATATCTACATTCATAAACA 1801
 XX
 RESULT 5
 ABL05640/C
 ID ABL05640 standard; CDNA; 4025 BP.
 XX
 AC ABL05640;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11402.
 XX
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 11402.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX
 OS Drosophila melanogaster.
 XX
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 P-PSDB; ABB61537.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 11402; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of

PT comprising the Casp8 gene that results in inactivation of the gene
 XX
 PS Example 3; Page 61; 107pp; English.
 CC Investigation of silencing of caspase-8 (CASP8) gene by methylation was
 CC analyzed using methylation-sensitive PCR analysis. Amplification of the
 CC 5' untranslated region (UTR) of the Casp8 gene was performed in reaction
 CC mixtures containing bisulfite treated DNA. Primers AAs1818-23 were
 CC designed to produce a 320 bp fragment in the upstream region of Casp8
 CC gene extending from nucleotides +221 to +541. Wild type primers were
 CC used to amplify the corresponding region of untreated genomic DNA.
 CC Controls without DNA were also performed. CASP8, a cysteine protease, is
 CC part of the death inducing signaling complex (DISC) associated with the
 CC Fas receptor. CASP8 is inactivated in cancers, and plays a role of a
 CC tumour suppressor gene. The CASP8 promoter region sequences, in
 CC particular Region 1 and Region 2, are crucial to the design and execution
 CC of the genomic methylation PCR analysis of CASP8 gene inactivation.
 CC Methylation PCR can be used to examine even minute amounts of patient
 CC material to demonstrate whether the CASP8 gene expresses an mRNA and
 CC protein product. The CASP8 gene has been localized to human chromosome
 CC 2q33-34. The methods are used to diagnose or prognose cancer. Cancer is
 CC treated by administering a vector that expresses a gene encoding
 CC functional CASP8 in cells. The cancer that is diagnosed or treated is a
 CC tumour in which a myc gene is amplified, such as a neuroblastoma.
 CC Aggressive neuroblastoma, juvenile neuroblastoma (preferred), small-cell
 CC lung carcinoma, non-small-cell lung carcinoma, colorectal carcinoma, or
 CC uterine cervical carcinoma can be diagnosed with the new method. A kit
 CC for screening for a compound that induces death-receptor-mediated
 CC apoptosis in cells containing an inactivated CASP8 gene is also
 CC provided.
 CC
 CC Sequence 25 BP; 12 A; 6 C; 0 G; 7 T; 0 other;
 SQ
 QY Query Match 100.0%; Score 25; DB 21; Length 25;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 CCATATATCTACATCAAAACAA 25
 1 CCATATATCTACATCAAAACAA 25
 RESULT 2
 AAs46772/c
 ID AAs46772 standard; DNA: 17280 BP.
 XX
 AC AAs46772;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Tumour suppressor gene derived chemically modified sequence #496.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200168912-A2.
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPig-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;

XX
 DR WPI: 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID No 496; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (Cp DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Se) and sequences complementary to (Se). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes. Sequences with even numbered Seq ID numbers are the
 CC complementary sequence of the corresponding odd numbered sequence (e.g.
 CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence
 CC is missing).
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 CC
 CC Sequence 17280 BP; 4730 A; 151 C; 3624 G; 8775 T; 0 other;
 SQ
 QY Query Match 80.8%; Score 20.2; DB 22; Length 17280;
 Best Local Similarity 88.0%; Pred. No. 88;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 DB 1 CCATATATCTACATCAAAACAA 25
 16705 CCAATATATCTACATTCACAAA 16681
 RESULT 3
 ABV50732/c
 ID ABV50732 standard; cDNA: 523 BP.
 XX
 AC ABV50732;
 XX
 DT 17-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 50723.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US05171.
 XX
 PR 17-FEB-2000; 2000US-183319P.
 PR 16-MAR-2000; 2000US-189862P.
 PR 25-MAY-2000; 2000US-207454P.
 PR 09-JUN-2000; 2000US-211314P.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 8, 2003, 00:03:20 ; Search time 24.9421 Seconds
(without alignments)
2257.229 Million cell updates/sec

Title: US-09-477-082-32

Perfect score: 25

Sequence: 1 ccatatatactacattcaaacaa 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues 4370478

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	AAA51821	Antisense primer f
2	20.2	80.8	17280	AA546772	Tumour suppressor
3	19.8	79.2	523	ABV50732	Human prostate exp
4	19.8	79.2	5474	ABO67066	Human angiogenesis
5	19.2	76.8	4025	AB105640	Drosophila melanog
6	19.2	76.8	4654	AA546781	Tumour suppressor
7	19.2	76.8	4654	AB134223	Human immune syste
8	19.2	76.8	8845	AA546544	Tumour suppressor
9	19.2	76.8	15373	AB132466	Human immune syste

C	10	18.8	75.2	5520	24	ABL33519	Human immune syste
C	11	18.8	75.2	6297	24	ABL32705	Human immune syste
C	12	18.8	75.2	9905	24	ABL32062	Human immune syste
C	13	18.6	74.4	221	22	AA161518	Soybean 240017 reg
C	14	18.6	74.4	361	22	AA191681	Human polynucleoti
C	15	18.6	74.4	379	22	AAK61674	Human immune/haema
C	16	18.6	74.4	798	24	ABK30312	Human G-protein-co
C	17	18.6	74.4	1463	19	AAV61483	Human secreted pro
C	18	18.6	74.4	1463	24	ABO92010	Human polynucleoti
C	19	18.6	74.4	1567	22	ABA08735	Human secreted pro
C	20	18.6	74.4	1602	21	AAZ52465	HTRM clone 1821233
C	21	18.6	74.4	4792	24	ABO54869	Human ovarian anti
C	22	18.6	74.4	5572	24	ABL54360	Chemically treated
C	23	18.6	74.4	5572	24	ABL53425	Human immune syste
C	24	18.6	74.4	6071	22	AA545373	Chemically pretrea
C	25	18.6	74.4	6071	24	ABK28210	DNA transcription
C	26	18.6	74.4	6960	22	AAK85579	Human immune/haema
C	27	18.6	74.4	7058	24	ABK40051	Human chemically p
C	28	18.6	74.4	10483	22	AAK80686	Human immune/haema
C	29	18.6	74.4	16236	24	ABL35023	Human immune syste
C	30	18.6	74.4	18011	24	ABL32034	Human immune syste
C	31	18.6	74.4	273254	21	AAK81914	Chlamydia pneumoni
C	32	18.6	74.4	335913	22	AA161371	Soybean 240017 reg
C	33	18.6	74.4	335913	22	AA161372	Soybean 240017 reg
C	34	18.6	74.4	1230025	20	AAK81990	Nucleotide sequenc
C	35	18.4	73.6	10279	24	ABL92276	Chemically treated
C	36	18.4	73.6	10279	24	ABL33590	Human immune syste
C	37	18.4	73.6	10279	24	AAK22327	Chemically treated
C	38	18.4	73.6	17294	24	ABL32987	Human immune syste
C	39	18.2	72.8	294	22	AA183128	Human polynucleoti
C	40	18.2	72.8	387	22	AA544850	Human cortig polyn
C	41	18.2	72.8	646	22	AAH87775	Peppermint plant o
C	42	18.2	72.8	672	23	ABL11307	Drosophila melanog
C	43	18.2	72.8	685	22	AA196618	Human neuroblastom
C	44	18.2	72.8	700	22	AAH92497	Human inflammatory
C	45	18.2	72.8	700	22	AAH92498	Human inflammatory

ALIGNMENTS

RESULT 1	AAA51821	standard; DNA; 25 BP.
ID	AAA51821	
AC	AAA51821	
XX		
DT	31-OCT-2000 (first entry)	
XX		
DE	Antisense primer for Casp8 unmethylated-specific disulfite treated DNA.	
XX		
KW	CASP8; caspase-8; Promoter; inactivation; methylation; cysteine protease;	
KW	tumour suppressor; chromosome 2q33-34; neuroblastoma; cancer;	
KW	death receptor; apoptosis; cytostatic; gene therapy; primer; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200039347-A1.	
XX		
PD	06-JUL-2000.	
XX		
PF	30-DEC-1999; 99WO-US31280.	
XX		
PR	31-DEC-1998; 98US-0114308.	
XX		
XX		
PA	(STUD-) ST JUDE CHILDREN'S RES HOSPITAL.	
XX		
PI	Kidd VJ, Lahti JM, Teltz T;	
XX		
DR	WPI; 2000-452423/39.	
XX		
PT	Detecting inactivation of a caspase-8 (CASP8) gene for diagnosing or	
PT	prognosing cancer, comprises detecting a modification of genomic DNA	

SOURCE
ORGANISM Homo sapiens.
REFERENCE Homo sapiens.
AUTHORS Mamalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 140137) Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE Waterston, R.H.
JOURNAL The sequence of Homo sapiens clone
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 140137)
TITLE Waterston, R.H.
JOURNAL Direct Submission
Submitted (30-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 16, 2000 this sequence version replaced gi:7770671.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0159F08
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Sequencing vector: M13; 100%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 128904 bases at least Q40
Consensus quality: 13055 bases at least Q30
Consensus quality: 135453 bases at least Q20
Insert size: 122000; agarose-ff
Insert size: 140532; sum-of-ctrls
Quality coverage: 3.10 in Q20 bases; agarose-ff
Quality coverage: 3.49 in Q20 bases; sum-of-ctrls

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
755 754: contig of 754 bp in length
855 854: gap of unknown length
1934 1934: contig of 1080 bp in length
1935 2034: gap of unknown length
2035 4109: contig of 2075 bp in length
4110 4209: gap of unknown length
4210 5692: contig of 1483 bp in length
5693 5792: gap of unknown length
5793 7261: contig of 1466 bp in length
7262 7361: gap of unknown length
7362 9539: contig of 2178 bp in length
9540 9639: gap of unknown length
9640 11725: contig of 2086 bp in length
11726 11825: gap of unknown length
11826 14512: contig of 2687 bp in length
14513 14612: gap of unknown length
14613 17420: contig of 2808 bp in length
17421 17520: gap of unknown length
17521 20684: contig of 3164 bp in length
20685 20784: gap of unknown length
20785 23973: contig of 3189 bp in length
23974 24073: gap of unknown length
24074 26703: contig of 2630 bp in length
26704 26803: gap of unknown length
26804 29324: contig of 2521 bp in length
29325 29425: gap of unknown length
29426 32886: contig of 3462 bp in length
32887 32986: gap of unknown length
32987 36388: contig of 3402 bp in length

FEATURES
SOURCE

misc_feature
36389 36488: gap of unknown length
36489 41815: contig of 5327 bp in length
41816 41915: gap of unknown length
41916 47324: contig of 5409 bp in length
47325 47424: gap of unknown length
47425 51087: contig of 3663 bp in length
51088 51187: gap of unknown length
51188 56711: contig of 5524 bp in length
56712 56811: gap of unknown length
56812 61865: contig of 5054 bp in length
61866 61965: gap of unknown length
61966 68716: contig of 6730 bp in length
68716 68815: gap of unknown length
68815 78915: contig of 10000 bp in length
78915 78916: gap of unknown length
78916 86751: contig of 7835 bp in length
86751 86850: gap of unknown length
86850 97390: contig of 10540 bp in length
97390 97490: gap of unknown length
97490 106694: contig of 9204 bp in length
106694 106795: gap of unknown length
106795 116444: contig of 9650 bp in length
116445 116544: gap of unknown length
116545 128169: contig of 11625 bp in length
128169 128269: gap of unknown length
128270 140137: contig of 11868 bp in length.

Location/Qualifiers
1..140137
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-159F8"
1..754
/note="assembly_name:Contig10"
855..1934
/note="assembly_name:Contig13"
2035..4109
/note="assembly_name:Contig16"
4210..5692
/note="assembly_name:Contig17"
5793..7261
/note="assembly_name:Contig19"
7362..9539
/note="assembly_name:Contig20"
9640..11725
/note="assembly_name:Contig21"
11826..14512
/note="assembly_name:Contig22"
14613..17420
/note="assembly_name:Contig23"
17521..20684
/note="assembly_name:Contig24"
20785..23973
/note="assembly_name:Contig25"
24074..26703
/note="assembly_name:Contig26
vector_end:17
26804..29324
/note="assembly_name:Contig27"
29425..32886
/note="assembly_name:Contig28"
32987..36388
/note="assembly_name:Contig29"
36489..41815
/note="assembly_name:Contig30"
41916..47324
/note="assembly_name:Contig31"
47425..51087
/note="assembly_name:Contig32"
51188..56711
/note="assembly_name:Contig33"
56812..61865

```

misc_feature complement(9232..9417)
      /note="match: GSS: Em:AQ062104"
      /note="match: STS: Em:G31247"
misc_feature 9243..9433
      /note="match: STS: Em:G31247"
misc_feature 9247..9358
      /note="match: STS: Em:G54157"
repeat_region 9433..9597
      /note="AluJo repeat: matches 132..292 of consensus"
repeat_region 9631..9904
      /note="AluJb repeat: matches 3..275 of consensus"
repeat_region 10052..10120
      /note="MLT2D repeat: matches 490..553 of consensus"
repeat_region 10142..10183
      /note="21 copies 2 mer ga 76% conserved"
repeat_region 10215..10651
      /note="MLT2D repeat: matches 1..498 of consensus"
repeat_region 10671..10960
      /note="AluSg repeat: matches 1..297 of consensus"
repeat_region 11084..11390
      /note="AluSx repeat: matches 1..304 of consensus"
repeat_region 11722..11767
      /note="L1P4 repeat: matches 5783..5828 of consensus"
misc_feature 11807..12289
      /note="match: GSS: Em:AQ358081"
repeat_region 11872..12194
      /note="AluSg repeat: matches 1..310 of consensus"
repeat_region 12355..12649
      /note="AluSp repeat: matches 1..302 of consensus"
repeat_region 12652..12960
      /note="AluY repeat: matches 1..308 of consensus"
repeat_region 13421..14475
      /note="MER32C repeat: matches 1..1162 of consensus"
repeat_region 14476..14781
      /note="AluSg repeat: matches 1..306 of consensus"
repeat_region 14782..14909
      /note="MER52C repeat: matches 1162..1278 of consensus"
repeat_region 15276..15580
      /note="MER52C repeat: matches 1162..1278 of consensus"

Query Match 80.8%; Score 20.2; DB 9; Length 25539;
Best Local Similarity 88.0%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAAACAA 25
Db 18759 CCAATGATCTACATCAAAACAA 18783

RESULT 14
AL365396 114016 bp DNA linear PRI 23-AUG-2001
LOCUS Human DNA sequence from clone RP11-338N12 on chromosome 9, complete
DEFINITION
ACCESSION AL365396
VERSION AL365396.10 GI:15384808
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 114016)
AUTHORS Wray, P.
TITLE Direct Submission
JOURNAL Submitted (23-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Aug 31, 2001 this sequence version replaced gi:13990096.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all

```

regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30): an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMBASE; Information on the WORMBASE database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr9> RP11-338N12 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES

```

source Location/Qualifiers
      1..114016
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="9"
      /clone="RP11-338N12"
      /clone_1lb="RPCT-11.2"
      41280..41334
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc_feature 42423..42805
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc_feature 43815..43851
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc_feature 51123..51163
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc_feature 51961..52029
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc_feature 94609..94626
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
misc_feature 109428..109461
      /note="Sequence from overlapping clone RP11-366H6 (AL449304). Assembly confirmed by restriction digest."
BASE COUNT 37485 a 19998 c 19402 g 37131 t
ORIGIN

Query Match 80.8%; Score 20.2; DB 9; Length 114016;
Best Local Similarity 88.0%; Pred. No. 2.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATATCTACATCAAAACAA 25
Db 99446 CCAATGATCTACATCAAAACAA 99470

RESULT 15
AC027611 140137 bp DNA linear HTG 07-JUL-2000
LOCUS Homo sapiens chromosome 12 clone RP11-155F8, WORKING DRAFT
DEFINITION
ACCESSION AC027611
VERSION AC027611.4 GI:8569832
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```


QY 4 TATATATCTACATTCATAACAA 25
 |||||
 Db 136196 TATATATCTACATTCATAACAA 136217

RESULT 11
 AC122828
 LOCUS
 DEFINITION Mus musculus chromosome UNK clone RP23-215B8, WORKING DRAFT
 AC122828
 AC122828.1 GI:21206486
 HTG: HTGS_PHASE1, HTGS_DRAFT.
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 270987)
 McPherson, J.D. and Waterston, R.H.
 The sequence of Mus musculus clone
 Unpublished
 2 (bases 1 to 270987)
 McPherson, J.D. and Waterston, R.H.
 Direct Submission
 Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA

COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@watson.wustl.edu
 Project Information
 Center project name: M.BA0215B08

----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 268408 bases at least Q40
 Consensus quality: 269116 bases at least Q20
 Consensus quality: 269568 bases at least Q20
 Insert size: 195000; agarose-fp
 Quality coverage: 18.59 in Q20 bases; agarose-fp
 Quality coverage: 10.33 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1472: contig of 1472 bp in length
 * 1473 1572: gap of unknown length
 * 1573 2810: contig of 1238 bp in length
 * 2811 2910: gap of unknown length
 * 2911 47630: contig of 44720 bp in length
 * 47631 47730: gap of unknown length
 * 47731 107746: contig of 60016 bp in length
 * 107747 107846: gap of unknown length
 * 107847 270987: contig of 163141 bp in length.

Location/Qualifiers
 1. 270987
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-215B8"
 misc_feature 1..1472

/note="assembly_name:Contig5"
 1573..2810
 /note="assembly_name:Contig6"
 2911..47630
 /note="assembly_name:Contig7"
 47731..107746
 /note="assembly_name:Contig8"
 107847..270987
 /note="assembly_name:Contig9"

BASE COUNT 80382 a 56456 c 54330 g 79404 t 415 others
 ORIGIN

Query Match 81.6%; Score 20.4; DB 2; Length 270987;
 Best Local Similarity 95.5%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TATATATCTACATTCATAACAA 25
 |||||
 Db 262929 TATATATCTACATTCATAACAA 262950

RESULT 12
 AX251530/c
 LOCUS AX251530 17280 bp DNA linear PAT 05-OCT-2001
 DEFINITION Sequence 498 from Patent WO0168912.
 ACCESSION AX251530
 VERSION AX251530.1 GI:15984953
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct.
 synthetic construct.
 artificial sequences.
 1 (bases 1 to 17280)
 Olek, A., Piepenbrock, C. and Berlin, K.
 Diagnosis of diseases associated with tumor suppressor genes and
 oncogenes
 Patent: WO 0168912-A 498 20-SEP-2001;
 Epigenomics AG (DE)
 Location/Qualifiers
 1..17280
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

FEATURES
 source
 1..17280
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 4730 a 151 c 3624 g 8775 t
 ORIGIN

Query Match 80.8%; Score 20.2; DB 6; Length 17280;
 Best Local Similarity 88.0%; Pred. No. 4.1e+02;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCATATATCTACATTCATAACAA 25
 |||||
 Db 16705 CCAATATATCTACATTCATAACAA 16681

RESULT 13
 AL161775
 LOCUS AL161775 25539 bp DNA linear PRI 25-APR-2001
 DEFINITION Human DNA sequence from clone RP11-264M3 on chromosome 13, complete
 sequence.
 ACCESSION AL161775
 VERSION AL161775.20 GI:13872274
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 25539)
 Lovell, J.
 Direct Submission
 Submitted (25-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 Requests: clonerequest@sanger.ac.uk
 On Apr 27, 2001 this sequence version replaced gi:13396429.

COMMENT

KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 226726)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 JOURNAL Mus musculus, clone RP23-28P10
 REFERENCE 2 (bases 1 to 226726)
 AUTHORS Unpublished

TITLE Direct Submission
 JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
 REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
 AUTHORS 3 (bases 1 to 226726)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N.,
 Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,T.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
 Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., MacDonald,P., Major,J., Marguis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trifoglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission
 JOURNAL Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
 REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Jun 26, 2002 this sequence version replaced g1:21536040.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: MIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center Project name: L20899
 Center clone name: 28_P_10

----- Summary Statistics
 Sequencing Vector: Plasmid; N/A; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 224168 bases at least Q40
 Consensus quality: 225210 bases at least Q30
 Consensus quality: 225616 bases at least Q20
 Insert size: 240000; agarose-gel
 Insert size: 225926; sum-of-contigs
 Quality coverage: 7.9 in Q20 bases; agarose-gel
 Quality coverage: 8.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 1496: contig of 1496 bp in length
 * 1497 1596: gap of 100 bp
 * 1597 1853: contig of 257 bp in length
 * 1854 1953: gap of 100 bp
 * 1954 5276: contig of 3323 bp in length
 * 5277 5376: gap of 100 bp
 * 5377 9084: contig of 3708 bp in length
 * 9085 9184: gap of 100 bp
 * 9185 25659: contig of 16475 bp in length
 * 25660 25759: gap of 100 bp
 * 25760 48397: contig of 22638 bp in length
 * 48398 48497: gap of 100 bp
 * 48498 144109: contig of 95612 bp in length
 * 144110 144209: gap of 100 bp
 * 144210 200142: contig of 55933 bp in length
 * 200143 200242: gap of 100 bp
 * 200243 226726: contig of 26484 bp in length.
 Location/Qualifiers
 1..226726
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP23-28P10"
 /clone_lib="RP23-28P10"
 /clone_name="28_P_10"

FEATURES

source
 misc-feature
 1..1496
 /note="assembly-fragment"
 clone_end:Sp6
 vector_side:left
 1597..1853
 /note="assembly-fragment"
 1954..5276
 /note="assembly-fragment"
 5377..9084
 /note="assembly-fragment"
 9185..25659
 /note="assembly-fragment"
 25760..48397
 /note="assembly-fragment"
 48498..144109
 /note="assembly-fragment"
 144210..200142
 /note="assembly-fragment"
 200243..226726
 /note="assembly-fragment"
 clone_end:17
 vector_side:right
 BASE COUNT 65243 a 45772 c 45704 g 69207 t 800 others
 ORIGIN

Query Match 81.6%; Score 20.4; DB 2; Length 226726;
 Best Local Similarity 95.5%; Pred. No. 2e+02; I; Indels 0; Gaps 0;
 Matches 21; Conservative 0; Mismatches 1;